A Measurement of Jet Shapes in Proton-Proton Collisions at 7.0 TeV Center-of-Mass Energy with the ATLAS Detector at the Large Hadron Collider

Thesis by

Zachary L. Marshall

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

10 June 2011

(Defended 9 November 2010)

Abstract

A study of jet shapes is presented using 300 nb^{-1} of proton-proton collision data collected at a center of mass energy of 7 TeV using the ATLAS detector at the Large Hadron Collider. The analysis includes jets with rapidity $|y| < 2.8$ and with calibrated transverse momentum $40 < p_T < 600$ GeV. Results are compared with several tuned Monte Carlo programs. Jets are found to be wider than predicted, although the evolution of the jet shape with transverse momentum is well described.

Contents

List of Figures

List of Tables

Chapter 1 Introduction

For almost forty years, the Standard Model of particle physics has held up to nearly all experimental tests, providing a consistent and accurate description of nature across a broad range of energies. From almost the beginning, it was clear that the Standard Model would eventually break down [1]. It cannot provide a complete picture above energies of about 1 TeV. The Large Hadron Collider (LHC), a proton-proton collider at CERN near Geneva, Switzerland, was designed to reach energies of $\sqrt{s}$ =14 TeV in order to finally probe this key energy range and, possibly, reveal information on some of the shortcomings of the Standard Model. ATLAS, one of the general purpose detectors at the LHC, will search this new regime for physics beyond the Standard Model.

At the LHC, quark- and gluon-scattering dominates the signal for most new physics searches by many orders of magnitude. Most models of new physics also include decay chains involving quarks and gluons in the final state. These particles are not themselves detected. Instead, they produce "jets," collimated collections of charged and neutral hadrons. Alone, these objects can provide critical tests of the Standard Model predictions in this new energy regime. They must also be well understood in order to reduce the Standard Model backgrounds to new physics searches and to measure many new physics processes.

One critical part of understanding jets lies in the characterization of their properties. Well-constructed jet "shape" variables can provide information about the detector performance, as well as showering and hadronization processes in the Standard Model. Other jet shapes may be useful to discriminate the background events from the signal of new physics processes. This thesis reports on some of the first jet shape measurements using the ATLAS detector at the LHC. It includes shapes for the highest energy jets ever studied in a particle collider.

Chapter 2 introduces the key pieces of the Standard Model from a historical, theoretical, and phenomenological point of view. Jets are defined precisely, many of their features are explained, and the concept of a jet shape is introduced along with several jet shape variables.

Chapter 3 describes the LHC machine. The critical beam parameters for the 2009 and 2010 runs, as well as the design parameters, are explained. The details of the ATLAS

detector relevant to this analysis are described in Chapter 4. Detector conditions for data taken during 2010 are described as well.

Chapter 5 presents the event selection used for the analysis of jet shape variables. The calorimeter noise suppression technique, jet calibration, and jet selection used by ATLAS are described. The method of unfolding from experimental results back to results comparable to theoretical quantities is explained. Systematic uncertainties are evaluated in detail for two of the historical jet shape variables. Chapter 6 then provides a comparison of jet shapes measured in 300 nb^{-1} of proton-proton collision data to the predictions of several Monte Carlo models. Jet shapes are compared in several difficult-to-describe regions of phase space, testing the models as completely as possible with the current data. Two new jet shape variables are introduced for the purpose of discriminating quark-jets from gluon-jets.

Finally, Chapter 7 presents the conclusions of the study. Some suggestions are made about future studies of jet shapes in ATLAS. A new unfolding technique is discussed which may reduce the systematic uncertainties of the measurement. Several interesting new variables are presented that might help to suppress Standard Model backgrounds in future searches for new physics. These variables are presented without a complete analysis of uncertainties, in order to hint at the possibilities that will be open once sufficient data have been collected.

Chapter 2 The Strong Interaction

The Standard Model has its origins in the quantum revolution of the 1920's and 1930's. Dirac's introduction of relativistic quantum mechanics in 1927 [2] made it possible to describe in mathematical detail the interactions between the fundamental particles in nature. At the time, there were only a few fundamental particles to describe: the proton, the electron, and the photon. The atom was fairly well-described as an electron cloud around a positively charged nucleus. One of the major open questions when using this picture was what force held together the protons inside the atomic nucleus. In order to overcome the electromagnetic repulsion, the force would have to be quite strong. At the same time, in order to not dramatically affect everyday life as we know it, it would have to have a very short range.

In 1934, Yukawa developed a theory of the strong force's interactions within the nucleus [3]. His theory included a force mediator, eventually identified as the pion, which had a mass of $\sim$100 MeV. The quantum was heavy enough to provide a sufficiently small range for the strong force, while appropriately binding the nucleus and not dramatically affecting nuclear β decay processes. The subsequent identification of the charged pion in the observation of cosmic rays in 1947 by Lattes, Occhialini, Powell, and Muirhead [4, 5, 6] seemed to confirm Yukawa's proposal. With the inclusion of anti-particles required by Dirac and discovered by Anderson in 1931 [7, 8], a complete picture of particle physics was coming together.

In 1947, Butler and Rochester published a photograph of a new particle in a cosmic ray shower [9]. This particle seemed to be neutral and to decay into two charged pions. It was later identified as the K_0. This discovery was the beginning of the end for the description of the proton as a fundamental object. Dozens more particles were discovered, first in cosmic ray showers, and then in bubble chamber experiments at early accelerators like the Cosmotron at Brookhaven and the Bevatron at Berkeley [10]. By the early 1960's, Gell-Mann and Ne'eman were attempting to classify all the known particles, and the classification scheme had become at least as complicated as Mendeleev's periodic table of the elements [11]. There were already extensive families of mesons (pion-like particles) and

baryons (proton-like particles).

Finally, a new model for dealing with these various particles was developed, called the "quark model" of particle physics [12, 13, 14]. The quark model explained the different observed particles as bound states of two or three fundamental particles, quarks. For example, protons were described as bound states of two up-quarks and one down-quark, neutrons as a bound state of two down-quarks and one up-quark. Each quark could carry a fractional charge, either $+\frac{2}{3}e$ (up-type quarks) or $-\frac{1}{3}e$ (down-type quarks), where $-e$ is the charge of the electron. The strong force in this model was explained by hypothesizing a force carrier, the "gluon," which dominated the internal dynamics of the hadrons (the particles composed of more than one quark). Proton scattering experiments at the Stanford Linear Accelerator Center (now the SLAC National Accelerator Laboratory) in the 1960's [15, 16] confirmed this picture of a proton as a composite object. Ultimately, these quarks were found to come in three different families (six flavors), each family including one up-type quark and one down-type quark. The last quark, the top, was found in 1995 at the Tevatron at Fermilab outside of Chicago [17, 18]. The quarks and gluons together are known as "partons," thanks in part to the early difficulties with the quark model.

This description of fundamental particles, however, was insufficient to explain what was *not* observed in nature. For example, no bare quarks or particles with fractional charge had ever been observed [19, 20], nor had any gluons been observed outside the nucleus. In order to solve this problem, it was suggested that quarks and gluons carry an additional "color" charge, analogous to the well-known electromagnetic charge [21, 22]. Quarks could carry one of three colors (typically green, red, or blue), anti-quarks one of three anti-colors, and gluons a color- and an anti-color charge. By postulating that only color singlet (i.e. colorless or neutral in color) states may exist in nature, only quark-anti-quark ($q\bar{q}$) states (identified as the mesons of Gell-Mann's picture) and three-quark states (identified as baryons) were allowed. Additionally, it is necessary to postulate that no gluon-only states, called "glueballs," may exist in nature [23, 24]. The latter is a purely *ad-hoc* assumption. Thus, the theory of the strong force came to be known as Quantum Chromodynamics (QCD).

The mathematical formalization of QCD was a team effort during the 1960's and 1970's. QCD was described as having an $SU(3)$ group symmetry that is physically interpreted as color symmetry. The collections of particles constructed in the early 1960's were revealed as

simply different representations of the $SU(3)$ group. This symmetry can be thought of as the group of unitary transformations acting on three-dimensional vectors (in this case, color charges), with one additional mathematical criterion: that the gluon color-singlet state may not exist in nature. The additional criterion results in many of the key features of the quark model, including, ultimately, asymptotic freedom. Asymptotic freedom means that at small distances the strong force becomes weak. The coupling constant of the strong force, α_s, is known to vary depending on the scale of the interaction (a variation known as "running"). At low energies (long distances), the coupling constant is quite strong, prohibiting perturbative calculations and forcing numerical solutions, as in lattice QCD [25]. At high energies (short distances), however, the strength of the coupling constant is sufficiently reduced to allow perturbative calculations. The concept of asymptotic freedom [26, 27] attempted to explain the range of the strong force as an effect of the running of the strong coupling constant. Eventually, this freedom was explained as a negative β function (the function that describes the scaling of the coupling constant of a field theory with energy) by Politzer, Gross, and Wilczek [28, 29]. The combination of QCD and the $U(1)_Y \otimes SU(2)_L$ symmetric electroweak force [30, 31, 32, 33, 34] finally forms the modern Standard Model [35]. With relatively few modifications, this theory has stood for almost forty years as the basis for all of modern particle physics.

At the LHC, protons collide at high energies, allowing for tests of the Standard Model and QCD in a new energy regime. Proton structure cannot yet be described from first principles[1], in large part because the strength of the strong coupling constant prohibits the use of perturbation theory. Instead, the proton's contents are characterized by "parton distribution functions" (PDFs) that describe the contribution of the fundamental particles to the proton's total momentum [36, 37, 38, 39]. These functions are typically written as the probability of finding a parton carrying a fraction of the proton's momentum x at a scale Q^2, which can be thought of as the distance scale being probed in the collision. An example of the proton PDF as a function of x at two values of Q^2 is shown in Figure 2.1 [36]. In the naive Quark Parton Model, the proton structure is well-described as two up-quarks and a down-quark. At high Q^2 and low x, the proton is a much more complex object, with significant contributions to its structure coming from sea quarks and gluons.

[1]Lattice QCD has made great progress in recent years towards describing the proton starting from the basic rules of QCD [25]. However, these results are not yet sufficiently advanced to provide a complete picture.

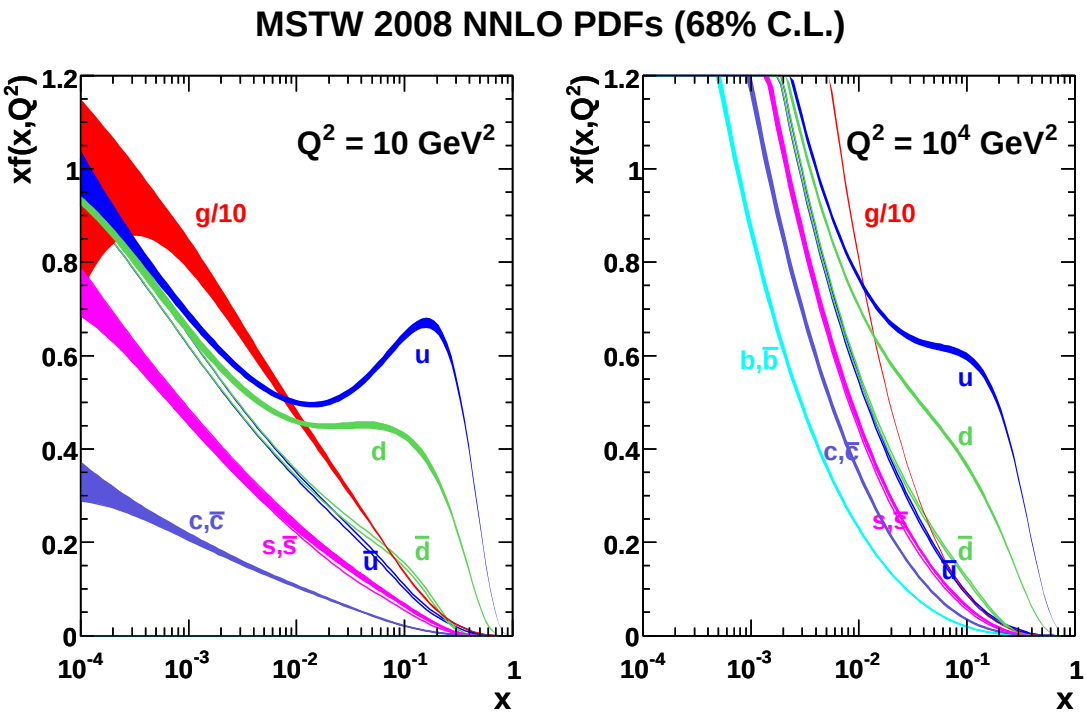


Figure 2.1: The parton distribution functions calculated by the MSTW collaboration of the proton at next-to-next-to-leading order (NNLO) for two values of Q^2. The shaded regions in the figures represent the 68% confidence level (C.L.) bands, the band that includes one standard deviation around the most probable value.

Generally, the PDFs are constructed using fits incorporating a great deal of experimental data, including $e-p$ scattering, neutrino scattering, muon scattering, and QCD measurements from the Tevatron. Since the LHC accesses regions of Q^2 and x that have not been measured, it is necessary to extrapolate the PDFs to ranges between and beyond those previously probed by experiments. The experimental results are compared to theoretical calculations using α_s at some order in perturbation theory (leading order (LO), next-to leading order (NLO), or next-to-next-to leading order (NNLO)), and the extrapolation is based on these calculations. As the low x and high Q^2 regime examined by the LHC is poorly constrained, different predictions may result from different collections of experimental data. These distributions will need to be modified and re-fit after the first LHC measurements become available, particularly to include the inclusive jet cross-section measurements.

In describing proton scattering, the PDFs must be used to determine, to first order, the momentum of the incoming partons. At the Tevatron, scattering cross sections for

most interesting processes were dominated by quark scattering, owing in large part to the significant contribution of quarks to the PDFs in proton-anti-proton scattering at a center-of-mass energy of 2 TeV. At the full energy of the LHC, however, gluon processes dominate. Thus, for example, top-quark production at the LHC is dominated by different processes than were dominant at the Tevatron. Additionally, the higher parton energy implies a larger open phase space for production of heavy particles like the W, Z, and top-quark.

For proton collisions at 14 TeV center of mass energy, the total cross section of ∼100 mb is dominated by elastic (pp final state) scattering and diffractive (pX final state) scattering. The total cross section is only known to about 25%. A surprisingly good estimate of the total cross section for proton-proton scattering comes from considering scattering using the classical radius of the proton, ~ 1 fm:

$$\sigma_{\text{total}} = 4\pi r^2 \approx 120 \text{ mb}$$

At the LHC, potential "new physics" beyond the Standard Model is expected to be suppressed by at least five orders of magnitude relative to 100 GeV jet production and by at least nine orders of magnitude relative to simple proton scattering. For many theories, these factors arise from avoiding the limits for new physics already set in previous experiments. The scaling of cross section with center-of-mass energy for several interesting processes can be found in Figure 2.2 [36]. The most interesting processes for the discovery of new physics have cross sections that are many orders of magnitude smaller than the total scattering cross section.

The simplest model of jet production in proton-proton scattering is simply a 2→2 QCD process, dominated by the Feynman diagrams shown in Figure 2.3 [40]. Gluon processes dominate because of color factors, but the PDFs are critical to selecting the correct combination of incoming partons. These high-energy processes, or "hard-scatters," result in the production of two outgoing final state partons and two beam remnants. At sufficiently high energies, the strong coupling constant is relatively small, and the processes can be fully described using perturbation theory. Nevertheless, higher order diagrams (i.e. diagrams with more powers of α_s) contribute a substantial cross section at the center-of-mass energy of the LHC [41]. Theoretical calculations for some processes exist up to NNLO, and numerical calculations at LO are not uncommon for as many as eight partons in the final state (see,

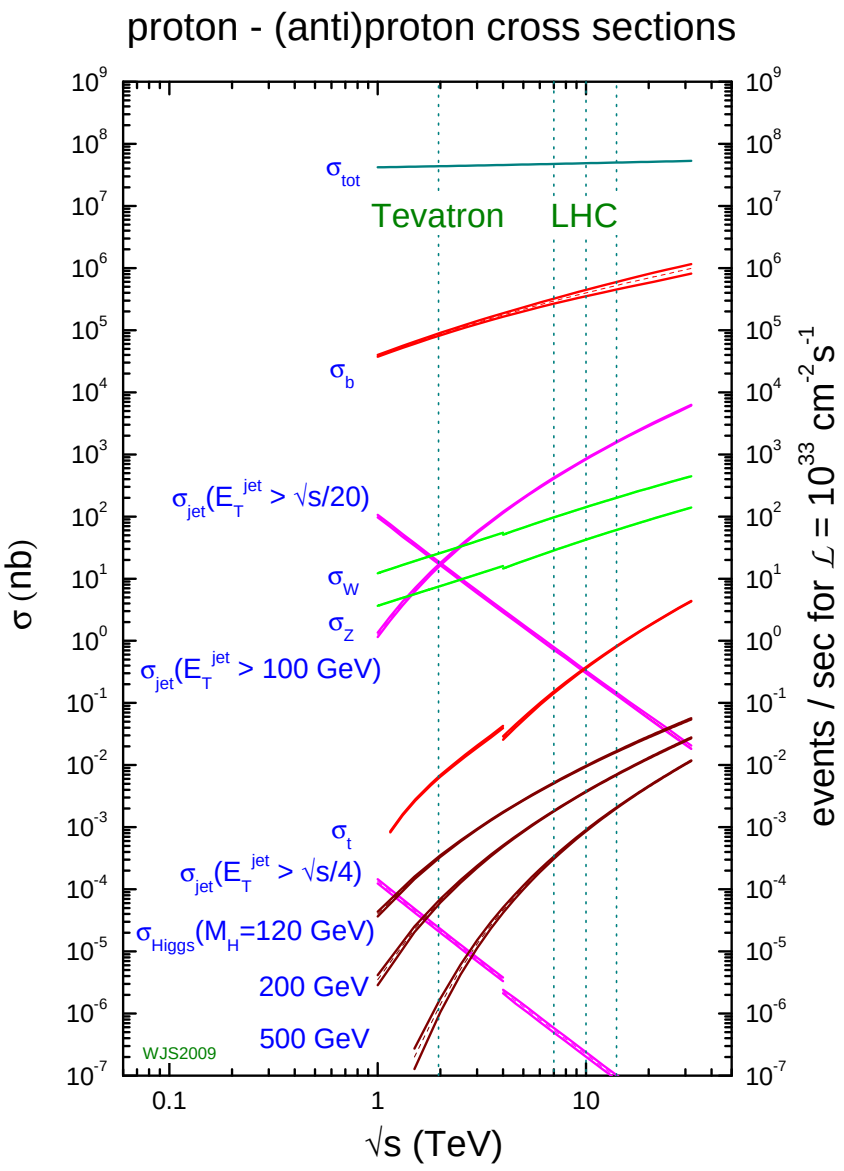


Figure 2.2: Cross sections for several interesting and proposed physics processes as a function of center-of-mass energy. The discontinuity at ∼4 TeV is from the transition from proton-anti-proton on the left to proton-proton collisions on the right. The vertical lines indicate the running energy of the Tevatron (1.96 TeV), the current running energy of the LHC (7 TeV), and two possible future running energies of the LHC (10 and 14 TeV). The cross sections shown are, from largest to smallest, the total scattering cross section ($\sigma_{\rm tot}$), the bottom-quark production cross section ($\sigma_{\rm b}$), the production cross sections for jets above several energy cut offs ($\sigma_{\rm jet}$), the W, Z, and top-quark production cross sections ($\sigma_{\rm W}$, $\sigma_{\rm Z}$, $\sigma_{\rm t}$, and the production cross sections for a theoretical Standard Model Higgs boson at with several masses ($\sigma_{\rm Higgs}$). The inclusive cross section is well described by gluon-exchange or the optical theorem, both of which are essentially independent of the differences between protons and anti-protons.

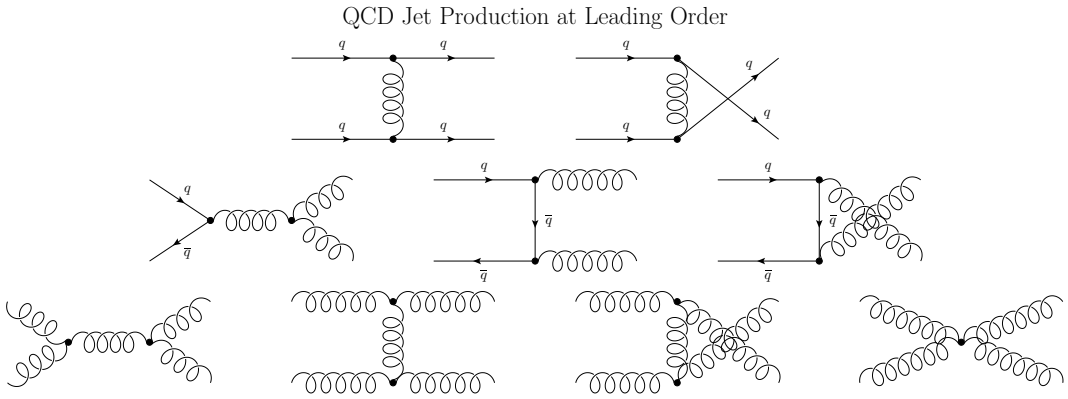


Figure 2.3: The dominant tree-level 2→2 Feynman diagrams for jet production in proton-proton scattering at the LHC. The top row shows quark-quark scattering diagrams, the second row shows quark-anti-quark scattering diagrams, and the bottom row shows gluon production diagrams. At the LHC, the top and bottom rows' diagrams dominate high p_T jet production. Gluon diagrams, of course, gain an extra factor from a larger color charge.

for example, Ref. [42]). Higher-order calculations have been limited by technical difficulties such as the huge number of Feynman diagrams involved and the removal of divergences [43]. The picture of proton-proton scattering as one of these hard processes is known as the "matrix element" picture. The term, of course, arises from writing a cross section as the squared sum of all amplitudes of the relevant Feynman diagrams (the matrix element itself).

It is important to keep in mind that the picture provided by all the calculations must be consistent for a fair comparison between theoretical and experimental results. That is, a PDF based on calculations at LO must be used with an LO matrix element, both of which must use an LO calculation of α_s. ATLAS is now using modified LO PDFs that may be able to "fake" NLO accuracy with only an LO matrix element, thanks to the recognition that many of the changes from LO to NLO are, in fact, due to PDF changes. Similarly, using an NLO α_s calculation with an LO PDF may improve agreement with experimental results. Both of these modifications result in a muddled theoretical picture. While they are useful for testing agreement between models and data, they cannot be used when testing agreement with QCD at any order.

This simple picture is, however, insufficient to fully describe the proton-proton scattering observed in experiments [44]. Several additional elements must be added to the gluon- or quark-scattering picture. High-energy radiation must be allowed from the incoming partons (called "initial state radiation") and in the outgoing partons (called "final state radiation").

The other partons in the proton must also be allowed to interact with one another (called "multi-parton interactions" - note that these are different from multiple proton-proton interactions in a single "event" at the LHC) [45]. The multi-parton interactions and initial-state radiation are energy-dependent and vary considerably from the energies of early colliders, to the Tevatron, and to the LHC. Generally, they are included with the beam remnant from the hard-scattering in the umbrella-term "underlying event," because none are a part of the matrix-element process. At the Tevatron, the underlying event was sufficiently similar to a typical non-single-diffractive inelastic ("minimum bias") event that the terms were used almost interchangeably [46]. Because the multi-parton interaction cross section grows with center-of-mass energy more quickly than the inclusive cross section, the differences become significant at the LHC and can no longer be neglected [47].

After the interaction, the outgoing partons "shower," producing many softer quarks and gluons. A parton shower may be used to approximate arbitrarily-higher-order processes without the computationally intensive calculations necessary to derive matrix elements for these processes. However, the description of the scattering as a $2\rightarrow2$ process is accurate so long as this radiation is much softer than the outgoing particles. As they approach the same scale, it becomes necessary to take into account higher-order processes at the matrix-element level.

It is important to only allow parton showers that would not otherwise have been produced directly in the hard scatter. For example, if one includes both $2\rightarrow2$ and $2\rightarrow3$ processes at the matrix-element level, allowing parton showers in the $2\rightarrow2$ events to produce similar event topologies to those described by the $2\rightarrow3$ matrix element would result in double-counting. In order to prevent this double-counting, two duplicate-removal schemes are commonly used. The first, called CKKW or CKKWL, vetoes any branchings in the parton shower that are more energetic than the cut-off for the matrix element production [48, 49]. Essentially, the parton shower for one process is never allowed to move an event from the original process (e.g. two jet production) into the phase space for another process (e.g. three jet production). The second, called MLM, re-clusters the partons after showering into a set of jets, and vetoes events based on the number of jets in the final state [50]. For example, $2\rightarrow3$ processes must create exactly three jets, and those three jets must match the three original partons reasonably well, or that event will be vetoed. If a parton shower is vetoed, the shower can be restarted from the matrix-element level without any penalty (except, of

course, for the time used to compute the unused shower).

Most of the showering process can be approximated by the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) evolution equations [51, 52, 53] that describe the probability for a quark to radiate a gluon or for a gluon to split into either a gluon pair or a quark pair. The showering process must be strictly ordered in some variable in order to ensure that the approximation is valid and that no double-counting occurs. Typical choices for this ordering variable are mass- (virtuality-) ordering [54], transverse momentum-ordering [54, 55] ("transverse" is the component of the momentum vector perpendicular to the incoming protons' momentum vector), and angular ordering [56, 57]. Angular ordering is accurate to the next-to-leading-log term, while transverse momentum ordering includes some higher order terms at the price of some errors in the leading-log contributions [58]. Ordering by virtuality fails to account for charge coherence in soft gluon emission (i.e. the fact that a soft gluon cannot distinguish nearby color charges).

Additionally, the parton showers must be rescaled to account for the probability of that shower having occurred [44]. This rescaling is done by the application of a "Sudakov form factor" that describes the probability for some radiation to *not* occur during the shower. In doing so, they ensure that virtual corrections and divergences are appropriately handled. In some sense, they correspond to a probability for that particular parton shower to have occurred, given the radiation phase space available to the initial partons in the event.

The last stage of the event is hadronization, where the partons coming from the shower are formed into stable, color-singlet hadrons. The energy scale for this process is generally around 1 GeV. This stage cannot be described by perturbative QCD and must therefore be described by some numerical model. There are two common models for hadronization. One, the Lund string model [59], produces hadrons by modeling (unphysical) "strings" stretched between colored partons. Conceptually, these strings arise from the dense strong-force field-lines between two colored objects. The string is allowed to break, and at each breaking point a quark-anti-quark pair is created. The string fragmentation continues until the only remaining particles are Standard Model resonances with insufficient string energy to break further. The other hadronization model, a cluster-based model [60], considers the endpoints of the parton showers to be massive "clusters" that "decay." These endpoints are required to be color singlets - any gluons are split into a pair of quarks prior to the beginning of the hadronization. The pairs of quarks (clusters) are split if high-mass, and

then the low-mass clusters are decayed into the known resonances of the Standard Model. Unlike the string model, only quark-anti-quark pairs may be hadronized in this way. No multi-parton clusters are allowed. Both of these hadronization models each involve several parameters that have been tuned to the available data. During hadronization, suppression of heavy (charm, bottom, or top) quark production is greater than 10^{10}, implying that it is very unlikely for heavy quarks to be introduced after the parton shower [44].

There are many implementations of both the parton shower and hadronization models used today. The most common are Pythia [54], Herwig [56, 57], and, to a somewhat lesser degree, Sherpa [55]. These programs are tuned to the available data [61, 62, 63, 64], but the tunes may not completely describe all regions of phase space [65, 66], particularly when lower-energy data are extrapolated to high center-of-mass energies. Generally, all the "steps" of the process described above are tuned simultaneously. For example, the hadronization and color-connection models are tuned with the inclusion of the underlying event. Turning off the underlying event would reduce the number of partons in the final state, thus changing the color connections in the event and the hadronization of the remaining partons.

2.1 Jets in Particle Physics

The stable hadrons coming from the proton-proton collision are the objects that may be observed experimentally. Some of them are short lived, and will decay before being directly detected. Eventually, most of the long-lived particles (except, of course, muons and neutrinos) will deposit their energy in the calorimeter of an experiment like ATLAS. In an attempt to resolve the partons coming from the hard scatter, these hadrons or energy deposits may be grouped into objects called "jets" [67]. These objects are the critical link between experiment and theory. Here, the discussion will focus on jets in proton colliders, but many of the theoretical and experimental challenges are very similar for other types of particle collider.

Because of the large production cross sections, jets are prodigious at the LHC. Already, jets are being used to "re-discover" the expected standard model processes and to ensure that the detectors are behaving as expected. Later, understanding jet productions rates and structure will be critical in controlling backgrounds to new physics searches. As shown

in Figure 2.2, the rate of production for jets over 100 GeV at a center-of-mass energy of 7 TeV is almost an order of magnitude higher than that of W production. W, Z, and $t\bar{t}$ production at the energies of the LHC often includes at least one jet from radiation or the underlying event, and the decay of the W, Z, or top-quark is most often hadronic (i.e. to jets).

Many different techniques have been developed for grouping objects into jets [67, 68]. In fact, a jet does not have a unique definition or existence independent of the algorithm that defines it. For example, a gluon that splits into a pair of quarks may be identified as two separate jets if the quarks are far enough apart and the jet cones are small, or as a single jet if the gluon is highly boosted or the jet cone is large[2]. Similarly, a boosted W that decays to a pair of quarks may appear as a single jet. Because of these features, there is no one-to-one correspondence between a parton at any stage of the shower and a jet in the final state. The mapping from partons to jets depends strongly on the jet algorithm.

Each jet algorithm has its own advantages and disadvantages, but they all share several common features. All algorithms collect some "constituents," inputs to the jet finding, into jets. These constituents may be experimental objects like calorimeter energy depositions or "tracks" (experimental traces left by charged particles), or theoretical objects like partons or hadrons. All algorithms also have some fundamental parameter, typically describing the size of the jet[3]. Small jets have the advantage of being less affected by particles from other interactions, including those coming from the underlying event. Large jets, on the other hand, are less affected by the spread of the jet due to parton showering, hadronization, and the showering of particles in the jet inside the experimental apparatus. Generally, small jets have been used for studies of events with high multiplicity final states (e.g. $t\bar{t}$ production [70]), and large jets have been used for low multiplicity final states (e.g. inclusive jet cross-section measurements [71]).

There are several characteristics of jet algorithms that are theoretically desirable [72]. The first is "co-linear stability:" the same jets should be reconstructed in an event if any constituent is replaced with two objects with the same momentum direction and half the original's energy. The second is "infra-red stability:" if two jets are produced nearby one another, some very soft radiation between the two jets should not cause them to be merged

[2]This example should be kept in mind during any discussion of a jet's partonic origin or quark "flavor."

[3]The variable-R cone algorithm is an exception to this rule [69]. This algorithm forms jets of a range of sizes. It is, however, difficult to calibrate experimentally and, therefore, has not been heavily used.

into a single jet. Of course, the jet algorithms should also be "easy" to implement in both experiment and theory. This is usually taken to mean that the computing time required to find the jets must be reasonable.

All jet algorithms cluster nearby objects in ϕ, the polar angle around the incoming protons, and rapidity, $y \equiv \frac{1}{2} \times \ln\left(\frac{E+p_z}{E-p_z}\right)$, where E is the energy of the object and p_z is its momentum along the direction of the incoming protons. For the construction of jets in "detector" coordinates, it is often convenient to use pseudo-rapidity, $\eta \equiv -\ln\left(\tan\left(\frac{\theta}{2}\right)\right)$, where θ is the azimuthal angle between the constituent and the incoming protons.

$$\begin{aligned} y &= \frac{1}{2} \times \ln\left(\frac{E+p_z}{E-p_z}\right) \\ &= -\frac{1}{2} \times \ln\left(\frac{E-p_z}{E+p_z}\right) \\ &= -\ln\left(\sqrt{\frac{E-p_z}{E+p_z}}\right) \end{aligned}$$

In the case of massless objects, where $E = |\vec{p}|$:

$$\begin{aligned} y &= -\ln\left(\sqrt{\frac{|\vec{p}|-p_z}{|\vec{p}|+p_z}}\right) \\ &= -\ln\left(\sqrt{\frac{1-p_z/|\vec{p}|}{1+p_z/|\vec{p}|}}\right) \\ &= -\ln\left(\sqrt{\frac{1-\cos(\theta)}{1+\cos(\theta)}}\right) \end{aligned}$$

Therefore,

$$y = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right) \equiv \eta$$

Rapidity has a more physical meaning for massive objects. For example, in the case of a two-parton final state, the average rapidity of the two outgoing partons is essentially the boost of the system relative to the lab frame. However, pseudo-rapidity is of particular use experimentally. Since pseudo-rapidity is the same for all objects (i.e. objects of any mass), it can be used in place of the azimuthal angle (θ) to define the coordinate system in a particle detector. For objects that have some physical meaning, like charged particles,

it is common to define the object as having the mass of a charged pion and to use rapidity to describe the object's direction. For objects that have no physical meaning, like cells in a calorimeter, it is conventional to define the object as having no mass and to use rapidity and pseudo-rapidity interchangeably.

After constituents have been grouped by the jet algorithm, their momenta must be combined to provide a single jet energy and momentum. The Snowmass recombination scheme uses the simple sum of the transverse energy of all constituents and the transverse energy-weighted direction [72] to create a massless jet. Alternatively, the constituents' momenta may be combined by summation of four-vectors. These two combination schemes will produce different jet masses, and may produce different jet structures [73].

The simplest algorithm, and historically most common, is an iterative seeded cone algorithm [72]. Some subset of constituents, typically those with a p_T above some threshold, are used as "seeds" to the algorithm. This algorithm constructs simple cones in $y - \phi$ space around these seeds, including all constituents within the radius R (where the distance $\Delta R \equiv \sqrt{\Delta\phi^2 + \Delta y^2}$) of the axis of the jet. The jet axis is adjusted to coincide with the energy- or transverse energy-weighted centroid of those constituents. Afterwards, a new list of jet constituents is constructed from those objects within R of the axis of the jet. This adjustment may be repeated several times, or until the position of the cone is stable. In this case, the Snowmass recombination scheme would produce a jet axis that points to the final centroid.

One disadvantage of this particular algorithm is that the iterative re-positioning of the jet axis may pull the jet away from an energetic constituent. Thus, high-energy constituents might not be included in any jet. As these constituents might have originated in the hard-scatter of the event, their having been missed can result in a poor reconstruction of the hard-scatter. Several methods have been proposed for reducing, but not eliminating, this effect. A smaller cone size can be used for the initial jet finding and centroid location than is used for the final collection of constituents. Alternately, the midpoints between all seeds can be included as additional seeds for the jet finding algorithm [74]. A third alternative is to remove the constituents included in a stable jet cone and begin the seeding process again (i.e.not allow any constituent to be included in two jets) [75, 76].

The cone algorithm requires an additional arbitrary parameter to treat overlapping jets. The parameter dictates when the two jets should be merged into a single jet and when they

should be "split" into two independent (nearby) jets. Such a treatment of overlap is not physically motivated.

Because of the requirement that some high-energy constituents "seed" the jets to be reconstructed, the cone algorithm is, by default, not collinear or infra-red safe. In order to improve the infra-red and collinear safety of cone jets, a modified cone algorithm called the Seedless, Infra-red Safe Cone (SISCone) algorithm has been proposed. SIScone does not use seeds like the traditional cone algorithm. From each set of constituents, a jet axis is calculated. That jet is kept if a cone around the axis includes exactly those constituents in the original set. It successfully identifies all possible cone jets, but the naive implementation requires computing time that grows exponentially with the number of constituents. A new implementation of this algorithm attempts to overcome the computational issues [77]. However, SISCone still suffers from the requirement of an additional jet splitting and merging stage, as well as still allowing the possibility of energetic constituents left outside of all final jets.

One alternative algorithm is the k_T algorithm [78, 79, 80]. This algorithm combines constituents that are nearby one another in momentum space. For each constituent i, a quantity $k_{T,i}$ is defined according to:

$$k_{T,i}^2 = p_{T,i}^2$$

where p_T is the component of the constituent's momentum transverse to the direction of the incoming proton. For every pair of constituents i and j, a quantity $k_{T,(i,j)}$ is defined according to:

$$k_{T,(i,j)}^2 = \min\left(p_{T,i}^2, p_{T,j}^2\right) \times \frac{\Delta R_{i,j}^2}{D^2}$$

where $\Delta R_{i,j} \equiv \sqrt{\Delta y^2 + \Delta\phi^2}$ is the opening angle between the two constituents in $y - \phi$ space, and D is the fundamental parameter of the jet algorithm related to the size of the final jets. If the $k_{T,(i,j)}^2$ for a pair of constituents is the smallest of all the $k_{T,i}^2$ and $k_{T,(i,j)}^2$, then the momenta of those two constituents are combined to form a new single constituent. If the smallest of all the $k_{T,i}^2$ and $k_{T,(i,j)}^2$ is instead the $k_{T,i}^2$ for a single constituent, then the constituent i is called a jet. The algorithm proceeds in this manner until all constituents have been combined into jets. Jets constructed with this algorithm are theoretically favored,

because they do a fairly good job of including all the p_T from the original partons (i.e. the mapping between partons and jets is good). Additionally, the k_T jet algorithm does not require an additional step and parameter to split or merge nearby jets. All splitting and merging is done by the jet algorithm itself during the jet finding. However, the jets have no well-defined area (i.e. they are not round like cone jets and may include objects far away in y and ϕ), and so suffer from many experimental difficulties.

A variation to this algorithm has been proposed [81], called the Anti-k_t jet algorithm. This algorithm substitutes a negative exponent in the fundamental equations for the k_T jet algorithm:

$$k_{T,i}^2 = p_{T,i}^{-2}$$

$$k_{T,(i,j)}^2 = \min\left(p_{T,i}^{-2}, p_{T,j}^{-2}\right) \times \frac{\Delta R_{i,j}^2}{D^2}$$

In doing so, the jet algorithm maintains many of the nice theoretical features of the k_T algorithm like infra-red safety and collinear safety, but the final jets constructed by the algorithm tend to be very conical. This jet algorithm is designed to avoid grouping many low-energy constituents into a jet. There are some indications that the Anti-k_t jets are constructed such that there is a minimum of energy at the edge of the jet [82]. That is, whereas cone algorithms operate by centering the jet on a local maximum of energy, Anti-k_t algorithms operate by centering the edge of the jet in a local minimum of energy. Similarly, where the k_T algorithm groups nearby constituents, the Anti-k_t algorithm tends to group high p_T objects together. It is important to remember that this difference in behavior is not "wrong," it is only a different feature of the jet finding algorithm.

One alternate variation on this algorithm is the so-called Cambridge/Aachen (C/A) jet algorithm [83, 84]. This jet algorithm uses a similar prescription as the k_T jet algorithm, but with:

$$k_{T,i}^2 = 1$$

$$k_{T,(i,j)}^2 = \frac{\Delta R_{i,j}^2}{D^2}$$

This algorithm is particularly interesting because it has no dependence whatever on the

p_T of the constituents. The k_T algorithm tends to group nearby, soft constituents together. This tendency can lead to grouping many soft constituents into high-energy jets that do not, in fact, approximate the momentum of any parton in the hard scatter. By combining constituents using only their proximity, the C/A algorithm reduces the number of these jets. In ATLAS, jets built with the C/A algorithm showed similar performance to those built using the Anti-k_t jet algorithm, with the exception that C/A jets had a broader distribution of jet area [85].

Each of these jet algorithms provides its own biases and features. Each also includes some parameter (R in the case of cone algorithms, D in the case of k_T-like algorithms) that characterizes the size of the jet. In order to ensure that no biases are introduced by the algorithmic definition of a jet, it is important to consider more than one jet-finding algorithm and more than one jet size. In fact, results with more than one jet size can be used to understand, for example, the relative importance of contributions due to underlying event (more significant for large jets) and hadronization (more significant for small jets).

2.2 Jet Shapes

Historically, jets have been measured using calorimeters, typically using the cone algorithm. This approach, however, omits a great deal of information available to the experimentalist and theorist. Jet substructure, for example the number of local maxima within a jet or the distribution of the jet energy across the cone, can provide useful information about the jet characteristics, parton showering, and hadronization. Similarly, information about the charged particle fraction of the jet, the fraction of jet p_T carried by tracks pointing towards the jet, can be used to glean further information about jet hadronization and particle decay within the jet. This information can prove critical for correctly measuring the energy of a jet [86, 87]. It can also provide information about the partonic origin of the jet that might be valuable in searches for exotic physics or highly boosted objects.

Jet shapes and jet moments are frequently used to characterize the properties of jets and have been measured in e^+e^-, ep, and hadron colliders [88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98]. Shapes may be built up from experimental quantities like calorimeter energy depositions or tracks, or from theoretical objects like hadrons or partons. The spatial granularity of the constituent object is typically negligible when a large enough sample of

jets is considered. For example, a jet may be made of only a few partons which shower into many hadrons. Shapes built from partons will inherently be more lumpy for a single jet than those built from the more numerous and lower-energy hadrons, but, over a sufficiently large sample, the *average* shapes should be comparable.

There is no need to define a jet's properties, shapes included, using the same "constituent" objects from which the jet was built. Famously, the discovery of the top-quark at the Tevatron used the information from tracks pointing towards jets that were built from calorimeter objects to identify jets originating from heavy quarks [17, 18]. In these cases, the jet algorithm is truly behaving as a jet "finder." Jet shapes and other characteristics, for example, may be built up from all objects within some radius of the jet, $\Delta R \equiv \sqrt{\Delta y^2 + \Delta\phi^2} < r$. In defining jet properties in this manner, one is free to look at the jet properties in an area within the jet or outside the area of the original jet. The extension of the shapes outside of the cone can be interesting for evaluation of the underlying event, the ability of the jet algorithm to locate the jet properly, and the physical "extent" of the jet, including its spread due to showering.

Ideally, jet shapes should be stable during the evolution from matrix-element level event through the parton shower and finally to stable hadrons in a particle detector. One example of this stability is also one of the simplest jet "shapes," jet mass. This "shape" is most commonly defined as the mass that results from the four-vector sum of all the jet's constituents. The mass of a jet may roughly correspond to the heavy particle that produced it, allowing for the identification of heavy or highly boosted particles [99]. Of course, there is some difference between the mass constructed from partons at the matrix-element level and the mass constructed from calorimeter towers, but the two are sufficiently similar that some boosted particle identification is possible.

Consider an annulus around a jet axis, from radius r to $r + \delta r$, where δr is some small value (typically 0.1), as shown in Figure 2.4. The sum of the transverse momenta of all constituents in this annulus can be written as:

$$A\left(r + \frac{\delta r}{2}\right) = \sum_{i: r < \Delta R_i < r + \delta r} p_T^i \tag{2.1}$$

A is centered in the annulus (i.e. at radius $r + \delta r/2$), p_T^i is the transverse momentum of constituent i, the sum is over all constituents, and $\Delta R_i \equiv \sqrt{(\phi_{\text{jet}} - \phi_i)^2 + (y_{\text{jet}} - y_i)^2)}$ is

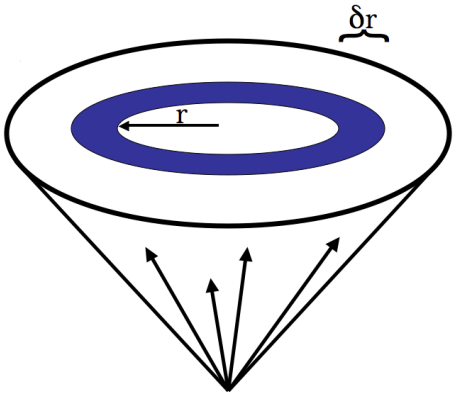


Figure 2.4: The shaded area shows a typical annulus within a conical jet with inner radius r and outer radius $r + \delta r$. As defined in the text, A is simply the sum of the p_T of all constituents within this shaded area. The jet shape variable $\rho\left(r + \delta r/2\right)$ is simply the average A over a large sample of jets. Typically, ρ is normalized to the sum of the p_T of all constituents within the cone of the jet (the outer circle in this diagram) and carries an additional factor of $1/\delta r$. ρ^a is simply ρ, normalized to the area of the annulus rather than by $1/\delta r$.

the opening angle between the jet and constituent i, where the jet (constituent) rapidity is y_{jet} (y_i) and the jet (constituent) polar angle is ϕ_{jet} (ϕ_i). Then one can write the differential jet shape:

$$\rho\left(r + \frac{\delta r}{2}\right) \equiv \frac{1}{\delta r} \times \frac{1}{N_{\text{jets}}} \times \sum_{\text{jets}} A\left(r + \frac{\delta r}{2}\right) \tag{2.2}$$

where the sum is over all N_{jets} jets in the sample, and the normalization by $1/\delta r$ is by convention. Because of the area dependence of this shape, it is convenient to instead define it per unit-area:

$$\rho^a\left(r + \frac{\delta r}{2}\right) = \frac{\delta r}{\pi\left[(r + \delta r/2)^2 - (r - \delta r/2)^2\right]} \times \rho\left(r + \frac{\delta r}{2}\right) \tag{2.3}$$

One might also consider the integrated jet shape:

$$\Psi\left(r\right) \equiv \frac{1}{N_{\text{jets}}} \times \sum_{\text{jets}} \sum_{i=\delta r/2}^{r-\delta r/2} A\left(i + \frac{\delta r}{2}\right) \tag{2.4}$$

This shape describes the transverse momentum contained in concentric circles around the axis of the jet, as shown in Figure 2.5. It is particularly convenient for describing the

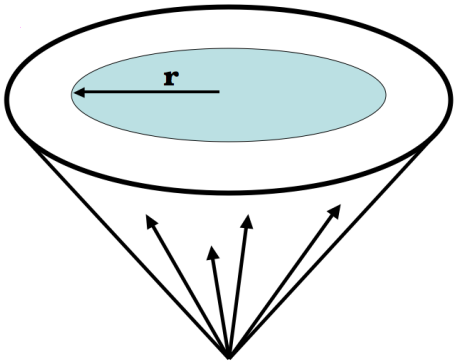


Figure 2.5: The shaded area shows a typical circle within a conical jet with radius r. The jet shape variable $\Psi(r)$ is simply the sum of the p_T of all constituents within this shaded area, averaged over a large sample of jets. Typically, Ψ is normalized to the sum of the p_T of all constituents within the cone of the jet (the outer circle in this diagram).

p_T- or y-dependence of the jet shape.

These shapes, $\Psi(r)$ and $\rho^a(r)$, will be the primary focus of this analysis.

The typical normalization for each of these shapes is based on the sum of elements "in the jet." The definition of "in the jet" can be either based on the algorithm (e.g. those elements which have been included in the jet by the jet finding algorithm) or based on geometrical separation (e.g. those elements which are within some distance of the jet in ΔR). Historically, it has been common to use algorithmic association with $R = 0.7$. In this case, both shapes are simply normalized to $\Psi(0.7)$.

These jet shapes can be used to demonstrate that a "jet" is not a uniquely defined object. The jet algorithm and jet size parameter significantly affect the properties of the jets. Figure 2.6 shows the jet shape $\rho^a(r)$ for several different jet algorithms constructed using stable hadrons in QCD di-jet events generated with PYTHIA [54]. All the jet algorithms shown here are described in Section 2.1. The k_T-like algorithms, Anti-k$_t$ [81], k_T [78, 79, 80], and Cambridge/Aachen k_T [83, 84], are all similar in shape when the jet size parameter is fixed. The similarity is somewhat surprising for the k_T algorithm in particular, which has an ill-defined jet area. When the jet size parameter is reduced, the jets appear significantly narrower. This is an expected effect of the transverse momentum requirement on the jets. For example, 60 GeV Anti-k$_t$ jet with size 0.6 has $\sim$15% of its transverse momentum in the outer-most two annuli. The core of the jet, therefore, would form a $\sim$50 GeV Anti-k$_t$ jet with size 0.4. The two cone-based jet algorithms shown [72, 74] have jet shapes that are

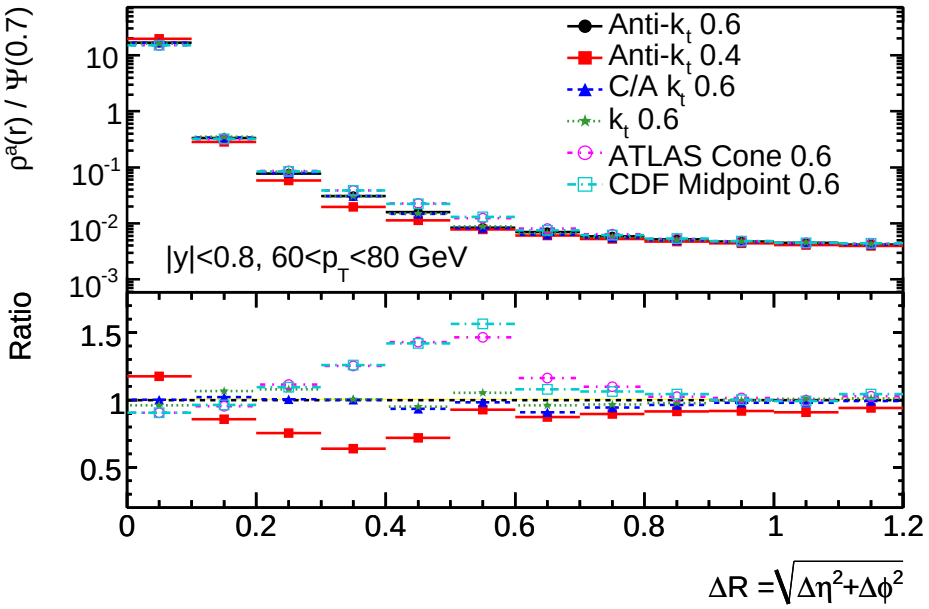


Figure 2.6: The jet shape $\rho^a(r)$ for several different jet algorithms. Generally, the k_T-like algorithms (k_T, Anti-k_t, and Cambridge/Aachen k_T) are similar in shape when the jet size parameter is held constant. The cone-line algorithms are considerably wider. Using a smaller jet size parameter produces a narrower jet shape, as expected.

wider by ∼50% at some radii. The difference results mostly from the splitting and merging of nearby jets. In a given transverse momentum range, merging will dominate because of the rapidly falling jet spectrum. Two merged jets of comparable transverse momentum will appear as a wide high-$p-T$ jet.

One can also define p_T moments of the jet:

$$M_\phi^1 = \frac{1}{N_{\text{jets}}} \times \sum_{\text{jets}} \frac{\sum_{i\in\text{jet}} p_T^i \times \Delta\phi_i}{\sum_{i\in\text{jet}} p_T^i} \tag{2.5}$$

$$M_y^1 = \frac{1}{N_{\text{jets}}} \times \sum_{\text{jets}} \frac{\sum_{i\in\text{jet}} p_T^i \times \Delta y_i}{\sum_{i\in\text{jet}} p_T^i} \tag{2.6}$$

$$M_\phi^2 = \frac{1}{N_{\text{jets}}} \times \sum_{\text{jets}} \frac{\sum_{i\in\text{jet}} p_T^i \times \Delta\phi_i^2}{\sum_{i\in\text{jet}} p_T^i} \tag{2.7}$$

$$M_y^2 = \frac{1}{N_{\text{jets}}} \times \sum_{\text{jets}} \frac{\sum_{i \in \text{jet}} p_T^i \times \Delta y_i^2}{\sum_{i \in \text{jet}} p_T^i} \tag{2.8}$$

where Δy_i is the opening angle between the constituent and the jet in y, $\Delta\phi_i$ is the opening angle between the constituent and the jet in ϕ, and again the sum is over jet constituents algorithmically or geometrically "inside" the jet. The first moments' mean and width provide a measure of the ability of a jet algorithm to center the jet on a local maximum of energy deposition and the corresponding position resolution inherent in the jet algorithm. For traditional cone algorithm jets with the Snowmass recombination scheme [72], the first moment should be near zero, depending on whether the constituents used to build the moment are the same as those used to build the jet.

The jet shapes ρ^a and Ψ and the second moments of the jets M_y^2 and M_ϕ^2 are, ultimately, all descriptions of a jet's width. The width of a jet is primarily a measure of the hard radiation in the event, as well as of the hadronization and fragmentation processes. In the case of the moments, since they are measured separately in y and ϕ, they can also provide measures of the symmetry of the jet[4].

As jets become more energetic, they narrow, and a larger fraction of their energy is contained in the core of the jet. At rest, decays are spherically symmetric. A very energetic jet is the result of a highly boosted parton, and special relativity dictates that the showering of that parton and decays of the subsequent hadrons will be more collimated as the boost increases.

Similarly, gluon-jets should be wider than comparably energetic quark-jets [96, 100, 101]. In the simplest picture of parton showers, all gluons decay into two quarks. Thus, a jet which has its origin in an energetic gluon should have a shape like that of two lower-energy (wider) quark-jets. One can alternatively understand the shape difference as an effect of the stronger color charge of the gluon, which leads to softer final-state particles [100]. The gluon radiation processes have an extra color factor relative to the quark radiation processes, $N_C/C_F = 9/4$. Thus, gluons are more likely to shower down into lower-energy particles,

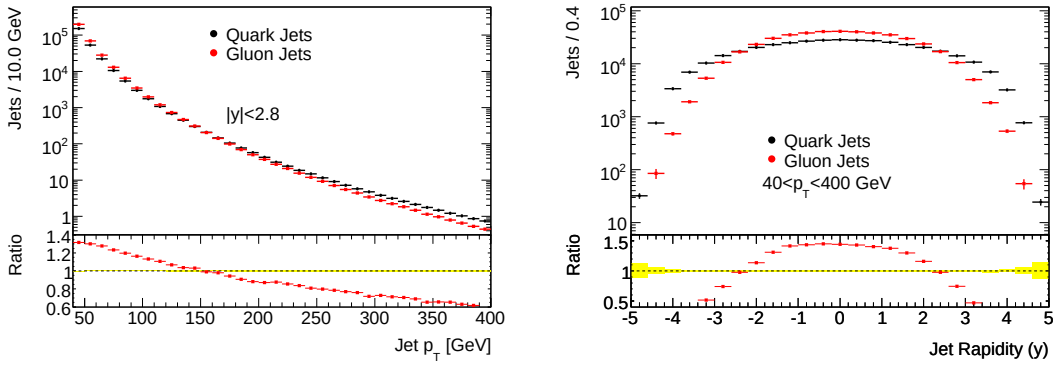
[4]The showering of the hadrons in the calorimeter is in cartesian (θ, ϕ) space, rather than in boosted (y, ϕ) space. Because the transformation from rapidity to θ is non-linear, a symmetric shower in (θ, ϕ) space will appear elliptical in (y, ϕ) space. Thus, an asymmetry between ϕ and y is introduced in the shower which must be accounted for in the comparison at the experimental level. The size of the effect increases with the rapidity of the shower.

and initial gluons produce broader showers with more low-energy particles. Therefore, these shapes provide a handle on the average quark-content or gluon-content of a sample of jets. Other shapes may provide some similar discrimination between quark-rich and gluon-rich samples of jets [102, 103].

It is critical at this point to remember the difference between distinguishing between quark- and gluon-jets *on average* and distinguishing between *individual* jets. Distributions and averages can be used to test, for example, parton distribution functions, but may not be usable with real confidence or significant efficiency to select the jet from radiation in a five-jet semi-leptonic $t\bar{t}$ event. It may be possible to build a weak discriminant from these shapes and other jet property variables, and such construction has been attempted at CLEO, LEP, and even in ATLAS [86, 100, 104, 105, 106, 107]. The task is somewhat simpler in an e^+e^- collider, where initial state hadronic radiation and the underlying event are negligible [108, 109]. Previous theoretically studies showed some promise at a hadron collider [101], but the discriminant's usefulness has never been fully demonstrated in practice.

Because of the different contributions to the proton PDF, it is expected that central jets in the LHC will be predominantly gluon-jets, and forward jets will be predominantly quark-jets. Similarly, low-energy jets predominantly originate from gluons at the LHC, and high-energy jets predominantly originate from quarks. This difference also reflects the differences in initial state partons involved in the collisions. The variation in jet shapes between the central and forward region and between low- and high-energy jets, therefore, can be used to extract some information about the PDFs or the quark or gluon content of a sample. Figure 2.7 shows the fraction of gluon-jets produced at a center-of-mass energy of 7 TeV as a function of the jet's transverse momentum and rapidity.

The OPAL experiment at LEP was able to study quark-jets and gluon-jets in highly enriched (>80% pure) samples and test their properties individually [107]. They found that gluon-jets have almost twice as many charged particles as do quark-jets, and that those particles are softer in the case of gluon-jets than in the case of quark-jets. All of their findings were consistent with the expectation based on color factors (2.25 as many charged particles in a high-energy gluon-jet as in a quark-jet). However, these studies were completed entirely in an electron-positron collider. Additional complications coming from color connections, beam remnants, initial-state radiation, and multi-parton interactions at hadron colliders make the creation of "pure" samples much more difficult.



(a) (b)

Figure 2.7: Gluon- and quark-jets produced at a center-of-mass energy of 7 TeV, as a function of the jet's transverse momentum (left) and rapidity (right). In large part because of the proton's PDF, it is expected that the gluon fraction will decrease at higher rapidity and at higher transverse momentum. The yellow band on the ratio represents the quark-jet distribution's statistical uncertainties. The ratio is defined as the number of gluon jets over the number of quark jets. The PDFs used to create these distributions were the modified LO PDFs [36].

Jets that include a bottom-quark are also expected to exhibit jet shapes different from those of jets originating from gluons, up-quarks, or down-quarks [110]. Generally, these jets are wider than light-quark-jets. Calorimeter jet shapes may assist in identifying these jets, but they are unlikely to provide a significant improvement over the current methods for identifying bottom-quark-jets. One particular experimental problem in tagging these jets is that for high energy gluons, the process $g \rightarrow b\bar{b}$ is expected to result in a single wide jet, rather than two separate b-jets [111]. These "$b\bar{b}$ jets" are markedly wider than jets that include only a single b-quark. In the analysis of the shapes of c-quark-jets, results have been obtained from photoproduction at HERA [112] and demonstrate the difference between these and light-quark-jets. Using these variables, one could construct a pure-calorimetric bottom-quark-jet tagging algorithm, albeit one far less efficient than the displaced vertex or semi-leptonic taggers normally used.

One more recent measure of jet size [113] uses the covariance tensor:

$$C = \frac{1}{\sum_{i \in \text{jet}} p_T^i} \times \sum_{i \in \text{jet}} p_T^i \times \Delta R_i \times \begin{pmatrix} \Delta y_i^2 & \Delta y_i \Delta \phi_i \\ \Delta \phi_i \Delta y_i & \Delta \phi_i^2 \end{pmatrix} \tag{2.9}$$

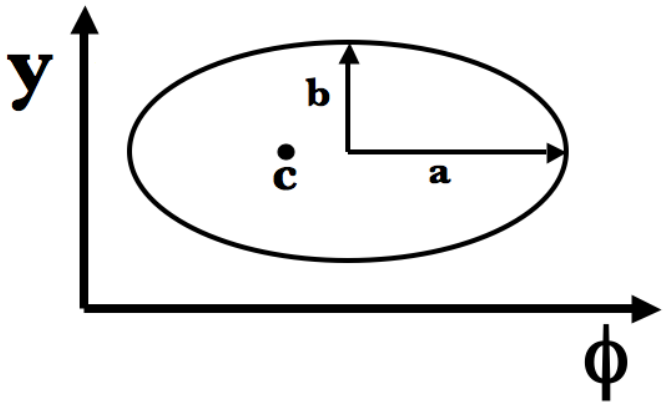


Figure 2.8: A diagram of an elliptical jet in $y - \phi$ space. The centroid of the jet is shown at point c. The major and minor axes of the ellipse have lengths a and b.

The eigenvalues of this tensor, a and b, can be thought of as the lengths of the major and minor axes of an ellipse around the jet (see Figure 2.8). These eigenvalues also allow one to define the ellipticity of the jet. For a jet ellipse with major axis in the $\hat{\eta}$ or $\hat{\phi}$ direction (as is the case in Figure 2.8), a and b are similar to the second moments in η and ϕ. The jet "size," $g \equiv \sqrt{a^2 + b^2}$, is similar to the jet area[5] and could be useful for distinguishing quark- and gluon-jets. Such a variable has been examined for the differentiation of quark- and gluon-jets using a slightly different form of the tensor [101].

More complex jet shapes can be constructed that may be valuable for new physics searches [114]. One example is jet "pull" [113], $\vec{p}$, which is defined as:

$$\vec{p} = \frac{1}{\sum_{i \in \text{jet}} p_T^i} \times \sum_{i \in \text{jet}} \left(p_T^i \times \Delta R_i \times \vec{r_i} \right) \tag{2.10}$$

where $\vec{r_i}$ is a two-dimensional vector in $y - \phi$ in the direction of the constituent from the jet axis and ΔR_i is the opening angle between the jet and constituent i. In the case of the jet in Figure 2.8, if there were sufficient energy deposited at large ϕ values, the pull vector would point in the ϕ direction from the center c. Because of the additional weight given to energy at large radius by the ΔR_i factor in Eq. 2.10, the energy deposition far from the center of the jet is most important for calculating the pull vector.

Both the magnitude and angle of the pull vector can be interesting for searching for the

[5] Note that this is *not* the same as the area of the ellipse defined by the eigenvalues of the tensor, $\pi \times a \times b$.

production of a colorless particle which may decay into colored particles, as in the case of the light Higgs boson decaying to two b-quarks. In this case, the two quarks coming from the decay must be in a color-anti-color pair, producing a color dipole. This color dipole would radiate just like the strong force analogue of an electromagnetic dipole. When the color dipole is boosted into the laboratory frame, the soft dipole radiation appears predominantly between the two quarks. If the two quarks are reconstructed as two independent jets, the pull vector of the first jet will point towards the second, and vice-versa.

In the case of QCD background to these decays, there is never a colorless particle produced. A simple picture for the case of quark scattering, for example, is that quarks with some color charge are "removed" from formerly colorless protons. The outgoing quark, which will be reconstructed as a jet, and proton remnant, which will normally continue down the beam pipe, form a color dipole. The radiation from this dipole, in the lab frame, will predominantly occur between the jet and the proton remnant. The pull vector constructed for the two jets in these events, then, will point towards the beam pipe, rather than towards any other jets in the event. This is, of course, a simplification of the true physics. Color re-connection during the parton shower and multi-parton interactions complicate the relationship considerably, washing out the contrast between the signal and background events.

Jet substructure can also be manipulated to improve the mapping from partons in the hard-scatter to jets in the final state. Two recently developed methods both involve attempting to remove soft contributions to an energetic jet that arise from the underlying event and initial-state radiation [115, 116]. At the LHC, and even during recent running at the Tevatron, it has been possible to have multiple proton-proton collisions in a single "event." These new methods may also help to remove background contributions to jets from other proton-proton collisions. Both methods may prove important in limiting, discovering, or exploring new physics at the LHC, particularly if the interesting final states are primarily hadronic.

All of these jet shapes may be defined by summing contributions from experimental or theoretical objects. In order to compare the shapes in experiment and theory, however, some "unfolding" must be done to remove different experimental effects that could alter the jet shapes. Some of these effects may mimic real fragmentation or decay processes. For example, any charged particles in the jet will bend in the magnetic field, removing some

particles from the cone of the jet algorithm and adding others that would not otherwise have been close enough to be included in the core of the jet. Because of the showering in the calorimeter, the jet will appear somewhat wider. The jet shapes may also be mis-measured because of experimental noise, background contributions (e.g. beam backgrounds or cosmic rays), or additional hard-scattering interactions. These effects, however, can all be studied with the implementation of a detailed detector simulation. After correcting the measured jets, it is possible to directly test different parton fragmentation models, showering models, parton distribution functions, and possibly even to measure the strong force coupling constant (though only to about 10% at the Z-pole) [117].

Chapter 3 The Large Hadron Collider

The Large Hadron Collider (LHC) [118, 119, 120] was constructed at CERN in the tunnel that housed the Large Electron-Positron (LEP) Collider [121, 122], a 27 kilometer circumference circular tunnel under France and Switzerland, just outside of Geneva. It was designed to collide protons every 25 ns at center-of-mass energies up to 14 TeV and lead ions at center-of-mass energies up to 5.5 TeV per nucleon-pair. This energy strikes a balance between exploration of new physics around 1 TeV and the superconducting magnet technology available at the time of construction [123].

Aside from the center-of-mass energy, the key parameter for a particle collider is "luminosity," measured in $\mathrm{cm}^{-2}\mathrm{s}^{-1}$, or events per second per unit of cross section. New physics processes likely appear far less frequently in proton-proton collisions than do the mundane processes of the Standard Model (see Figure 2.2). The luminosity of the LHC, therefore, must be quite high in order to fully explore this new energy regime. The luminosity of any collider is given by [118, 119, 120]:

$$L = \frac{n_b \times n_p^2 \times f \times \gamma}{4 \times \pi \times \epsilon_n \times \beta^*} \times F \tag{3.1}$$

where n_b is the number of bunches in the machine, n_p is the number of protons per bunch (symmetric between the clockwise and anti-clockwise beams), f is the frequency with which the collisions occur, γ is the relativistic gamma factor for the protons, F is a geometric factor due to the crossing angle of the beams, and ϵ_n (the normalized transverse beam emittance) and β^* (the β function of the beam at the collision point) are measures of beam spread in position and momentum space. The luminosity and cross section for a physics process give the rate of interactions at the collider. Thus, data set size is commonly measured in "integrated luminosity," with units of inverse cross section. Anti-protons are very difficult to produce and limited the luminosity at the Tevatron [124]. In order to reach the desired luminosity of 10^{34} cm^{-2} s^{-1}, the choice was made to use two proton beams at the LHC rather than a proton and an anti-proton beam. During 2009 and the first half of 2010, the LHC luminosity increased from $\sim 10^{26}$ to $\sim 10^{30}$ cm^{-2} s^{-1}.

To ensure collisions occur at the geometric center of the detector, and because of the difficulties associated with accelerating, controlling, and monitoring a continuous beam of protons, the LHC proton beam is in fact made up of a sequence of "bunches" of protons. The beam comprises a maximum of 2808 bunches of protons, each containing up to 1.15×10^{11} protons with an "abort gap" (empty space in the beam) of at least 3 μs. The design luminosity of the LHC is 10^{34} cm^{-2}s^{-1}, which results in a mean of 23 proton-proton collisions per bunch-crossing. At that luminosity, after $\sim 10 - 15$ hours the luminosity has dropped sufficiently to make it necessary to dump the beams and re-inject. During initial running and machine commissioning, the luminosity and all of the beam parameters are varied considerably. The parameters for the data collected in 2009 and the first half of 2010 are listed in Table 3.1, along with the design parameters. The optics of the LHC main ring are optimal at 7 TeV per beam. Below that energy, only lower luminosities are achievable.

Luminosity [cm^{2}s^{-1}]	Design (10^{34})	2009 Runs	2010 Runs
Center-of-mass energy [GeV]	14000	900-2360	900-7000
Number of bunches per beam	2808	1-9	1-8
Number of protons	1.5×10^{11}	5×10^{8}-5×10^{10}	10^{10}-10^{11}
Revolution frequency	11.2×10^{3}	11.2×10^{3}	11.2×10^{3}
Relativistic γ factor	7460	480-1260	480-3740
ϵ_n [μm rad]	3.75	2-4	1.5-10
β^* [m]	0.55	11	2-11
Geometric crossing factor	0.836	$\sim$1	$\sim$1
Instantaneous Luminosity [cm^{-2}s^{-1}]	10^{34}	$\sim$1-3$\times10^{26}$	$\sim 10^{27}$-10^{30}
Mean interactions per crossing	23	0	$\sim$ 0-2.2

Table 3.1: LHC operating parameters for design operating conditions and the runs analyzed here.

The accelerator complex, shown in Figure 3.1, comprises several stages that feed the main ring of the LHC [120]. Linac-2, a linear accelerator built in 1978 at CERN, produces a proton beam from a duoplasmatron source and boosts the protons to about 50 MeV. The linac can produce $\sim$120 μs-long bunches of protons with a total current over 180 mA. The bunches of protons are injected into the Proton Synchrotron Booster (PSB), where they are split into four parallel, identical rings, boosted to 1.4 GeV, and recombined into a single bunch. Each bunch is kept together with a radio frequency (RF) acceleration system, where the bunch tends to sit at the minimum of the RF (in a "bucket"). The Proton Synchrotron (PS) can hold up to seven bunches of protons from the PSB.

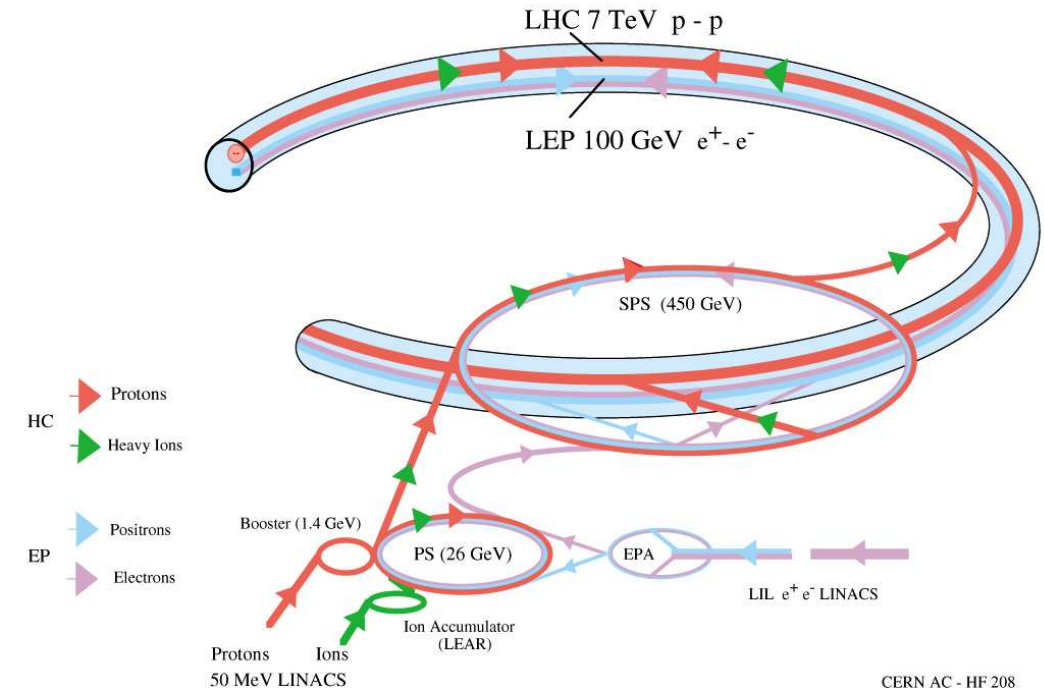


Figure 3.1: The injection complex and outline of the Large Hadron Collider. The linear accelerator for LEP (LIL) and electron positron accumulator (EPA) are shown in the bottom right corner and could potentially be used for a large proton-electron collider in the tunnel of the LHC.

For the remainder of the injection sequence, at least one bucket is always left empty (e.g. only six buckets are filled in the PS) in order to allow safe transfer of the protons into or out of each ring. To transfer the proton bunches from one ring to the next, a set of dipole magnets that can quickly turn on (called "fast kicker magnets") are included in the rings. These magnets turn on during the gap in the beam and cleanly deflect all the bunches of protons out of the ring and into a transfer line. This "abort gap" ensures that no proton bunch passes through the fast kicker magnets while they are turning on. Without it, at least one bunch of protons would be deflected between the ring and the transfer line, causing a radiation hazard and possible instrument damage.

During LHC operation at design luminosity, the PS will split the bunches in three after injection, accelerate the protons to 25 GeV, and, afterwards, split the new bunches in four. At this point the one empty bucket has become 12 and the six proton bunches have become

72. The proton bunches have $\sim 10^{11}$ protons in them and will not be split or combined any further. This design splitting was successfully achieved prior to 2009, although during the commissioning of the LHC, frequently fewer bunch-splittings were induced. Next, the protons are injected into the Super-Proton Synchrotron (SPS), the accelerator where the W and Z were first discovered. Up to four batches of 72 proton bunches can be injected into the SPS at a single time. There, the protons are boosted to 450 GeV and are finally injected into the main ring of the LHC. At design luminosity, this entire process is repeated 24 times over about 15 minutes in order to completely fill both beams in the LHC ring. During commissioning in 2009, no more than 16 bunches were injected into the LHC at a time, and the injection process took less than five minutes. After injection at 450 GeV, it would take approximately 20 minutes in the LHC to accelerate the beams to 7 TeV. This ramp time is limited mostly by the time needed to increase the current in the magnet system, so reaching a lower energy requires less time. Until August 2010, it took 30-45 minutes to ramp the beams to 3.5 TeV because of protection system problems limiting the maximum change in current in the LHC magnets, $dI/dt < 2$ A/s.

To save space, the main LHC ring transports both proton beams inside a single structure with a double-beam pipe, mostly defined by (1232) 15 m long, 40 tonne, twin-bore dipole (beam-bending) magnets. Over 3500 higher-order corrector magnets are installed around the ring to provide orbital stability and keep the beam focused. In order to allow the highest possible beam energy, requiring the strongest possible magnets, the majority of the magnets in the LHC ring are superconducting. A superfluid helium-II cooling system maintains their temperature at 1.9 K. Superfluid helium was chosen for its excellent heat transport properties, and the temperature was selected to allow some heating (a temperature rise of up to 0.7 K) without a phase transition in the helium.

ATLAS uses a right-handed coordinate system. The y-axis is vertical (up is positive), the x-axis points into the LHC ring, and the z-axis points roughly east (counter-clockwise). A cylindrical coordinate system is most convenient, in which R is defined as $\sqrt{x^2 + y^2}$ and ϕ is the azimuthal angle around the z-axis, with $\phi = 0$ corresponding to the x-axis. For detector coordinates, η will be used instead of polar angle (θ).

The luminous region in ATLAS during 7 TeV center-of-mass energy operation was well centered (e.g. (-0.370,0.628,1.0) mm in (x,y,z) with a fitted Gaussian length in z of $\sim$22.1 mm and a fitted Gaussian width of ~ 120 μm in x and ~ 132 μm in y for one of the

first fills at 7 TeV) [125]. The beam position in x and y is stable between fills to $\sim$100 μm, and the position in z is stable to a few mm. Luminosity scans, during which the beams were adjusted vertically and horizontally in order to center them on each other, were conducted once every few successful injections during the commissioning period.

The beam lifetimes, the time it takes for the current of the beam to decrease by $1/e$, during collisions are expected to be 10-15 hours, limited mostly by the collisions themselves [118]. The luminosity decreases over the lifetime of the beam because of both a reduction in the number of protons per bunch and an increase in the beam emittance. During the runs in 2009 and 2010, beam lifetimes over 1000 hours were achieved while the beams were not colliding or were colliding at low luminosity. Luminosity lifetimes, the time it takes for the luminosity to decrease by $1/e$, were 50-100 hours.

Several changes can be made to the beams after injection into the LHC so as to achieve a higher collision rate or longer luminosity lifetime. In order to increase the collision rate, the beams are "squeezed" in the transverse direction after the ramp is complete. In order to maintain better control of the beams, at high energy (i.e. above a few TeV in the main LHC ring) the proton bunches are artificially "blown up" in the longitudinal direction. This "blow up" roughly doubles the bunch length, thereby reducing intra-bunch interactions that can cause significant losses. Even with the longitudinal blow-up, these beams provide interactions very close to the geometric center of ATLAS. Once the beams are at full energy and no further adjustments need to be made, they are declared "stable" by the machine operator.

The LHC provides proton-proton collisions to six experiments at four different interaction points around the ring [126, 127], as shown in Figure 3.2. It is divided into eight sectors, between which are access points numbered clockwise around the ring, beginning with point 1 at the south-most point of the ring, nearest to the main site at CERN. The beams are injected from the SPS between points 1 and 2 and between points 1 and 8, and they are only crossed for collisions at four points: point 1, point 2, point 5, and point 8. ATLAS [128] sits at point 1, as does the forward scattering experiment LHCf [129]. Point 2 contains ALICE, an experiment designed mostly to measure heavy-ion collisions [130]. The CMS (Compact Muon Solenoid) general purpose detector [131] and TOTEM forward scattering experiment [132] are both located at point 5. LHCb, a B physics experiment, is located at point 8 [133]. LHCb is an asymmetric detector, and collisions occur on one side

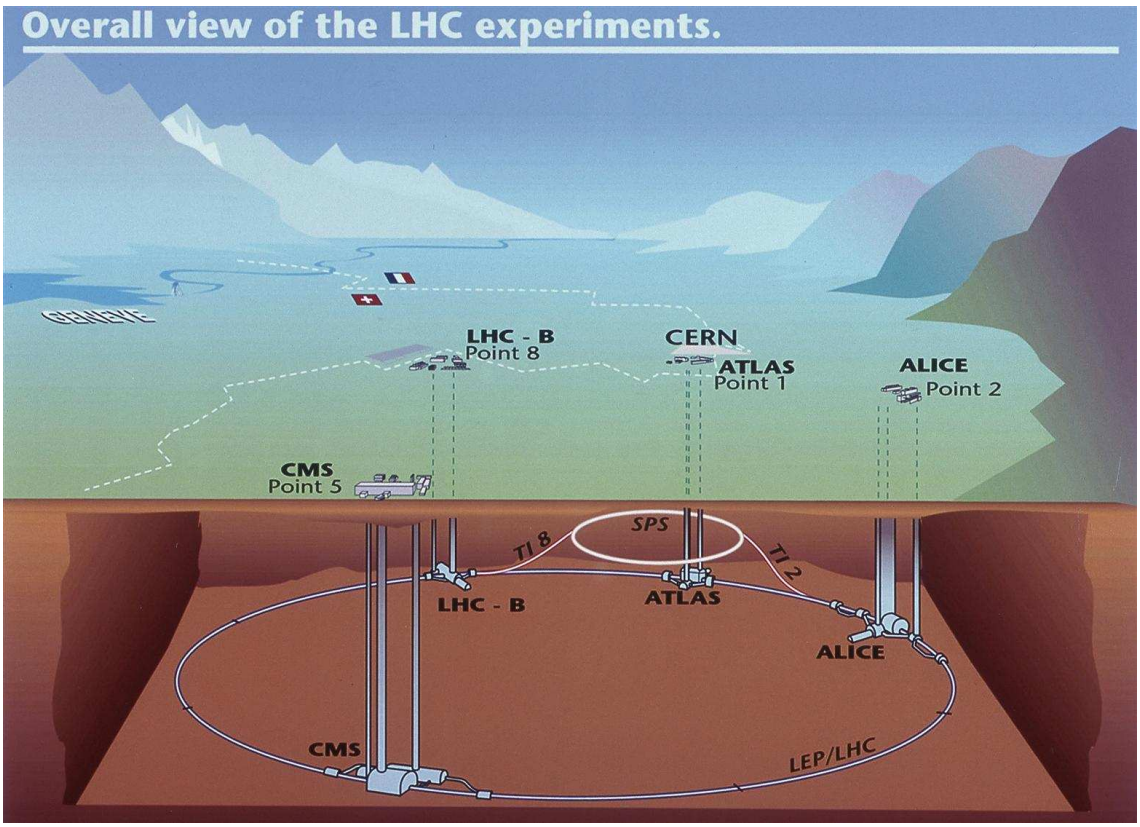


Figure 3.2: The interaction points housing the experiments at the Large Hadron Collider. ATLAS sits at point 1, toward the top of the image.

of the interaction region. If collisions are to occur in all four interaction points during a single fill, this asymmetry results in a strict limitation on the bunch spacing in the machine.

At the interaction region near point 1, the two proton beams are housed in a single beam pipe for 280 m (140 m on either side of the interaction point). Within that space, the inner triplet magnets, also called "final focus magnets," squeeze the beam transversely to focus it at the interaction point. Thanks to this squeeze, the beam bunches to be transported around the ring may be considerably larger than they need be at the interaction point, reducing intra-beam interactions. In initial operation with a limited number of bunches in each beam, there is no need to cross the beams at an angle. At design luminosity, however, the beams are crossed at an angle to prevent parasitic collisions from occurring every few meters throughout the experimental cavern. At point 1, for example, the beams will be crossed in the vertical plane, at an angle of $\sim$142 μrad.

Point 4 houses the RF acceleration system, used to accelerate the beams from 450 GeV

to 7 TeV. The RF system provides up to 16 MV at 400 MHz during beam operation at 7 TeV. The high frequency is required to maintain the beam's bunch structure. Because the LHC is a proton collider, energy losses from synchrotron radiation are small. For 7 TeV protons, around 7 keV is lost per turn or $\sim 10^{-9}$ of the proton's energy, compared to $\sim 2\%$ for LEP at around 200 GeV. The energy of the beam is very well measured. During the operation of LEP and the commissioning of the LHC, small deviations in beam energy were observed from the variation in ring size arising from the lunar tides [134].

Points 3 and 7 are used for beam cleaning, including collimation to maintain the stability of the beam. The collimators in point 3 are designed to scatter particles with deviant momenta, and collimators in point 7 scatter particles with large betatron amplitudes. Because of this cleaning and because of the difference in focus magnets in points 8 and 2, the counter-clockwise proton beam is expected to be almost 50 times cleaner when it arrives at ATLAS than the clockwise beam.

In the case of an operational problem, or at the end of the beams' lifetime during normal operation, the beams are dumped in point 6. During the beam abort gap, fast kicker magnets ramp on and deflect the beam, ultimately leading them into straight tunnels offset from the LHC ring. As they pass into the beam dump tunnel, the beams are defocused before they finally arrive on graphite core beam stops. In order to reduce the energy deposit in any single spot, the bunches of the beam are "painted" across the front of the graphite core, as shown in Figure 3.3. This spread allows safe stoppage of the beam, which will reach 360 MJ at design luminosity [118]. The clockwise and counter-clockwise beams are sufficiently far apart at point 6 that the beams can be dumped independently of one another.

Several incidents leading up to operation in 2009 and 2010 brought to light critical failure modes of the LHC. In order to reach the high magnetic fields required for operation, most of the magnets must be superconducting. Superconducting magnets can, however, undergo a "quench," during which, usually due to resistance from imperfections or junctions causing a temperature rise, the magnet ceases to be superconducting. Such quenches are expected as part of commissioning the magnet system of the LHC, but under some circumstances they can be problematic. In the case of the LHC magnets, many "training quenches" have been necessary to achieve maximum current. These quenches occur as the magnets reach a new peak current. The reason that the training quenches are necessary is not fully understood, but several thousand of them may occur before the machine is able to reach

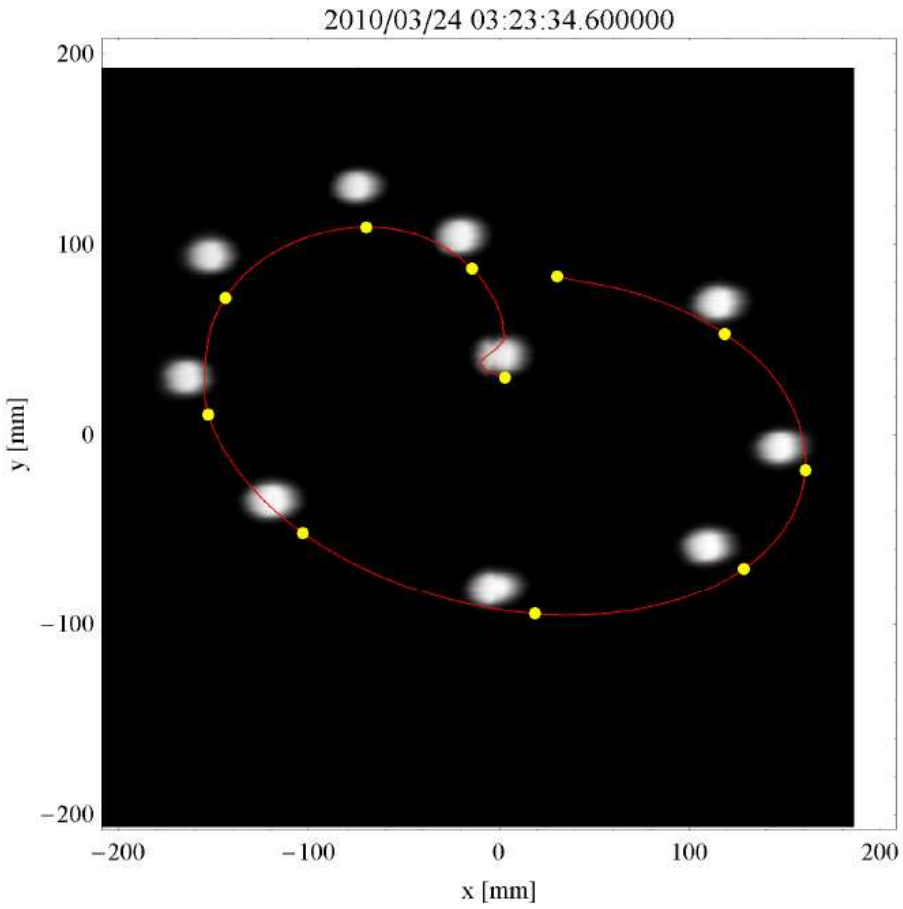


Figure 3.3: The beam dump at the LHC, with ten bunches dumped. Shown are the positions of the bunches as they are dumped (white spots), the curve showing the predicted distribution (red line), and the calculated positions of the bunches (yellow spots).

design energy [135].

During 2007, a quench of the inner triplet magnets near one of the interaction points halted operation for several months. The quench caused a significant force along the beam line, a direction in which little bracing had been included in the design. The magnet shifted inside its cryostat but did not cause damage beyond the immediate area around it. Subsequently, the inner triplet magnets were all reinforced prior to operation in 2008.

During 2008, in the final commissioning of the last remaining sector of the machine, a serious quench occurred, stopping operation for a year of repairs and redesign. Between each pair of magnets there is a joint where two superconducting wires meet. The two wires are laid one on top of the other, and copper bars squeeze them together, with solder sealing all the joints. In several cases, little or no solder was, in fact, in the joint. Because of this,

some resistance was present in one of the joints (at the several hundred nΩ level), causing a temperature rise in the joint and additional resistance. Because sensors throughout the machine were best suited to detect resistance within a single magnet, the joints between the magnets were particularly susceptible to problems. During high-current tests, the voltage across the joint built up sufficiently that an electrical arc formed, burning a hole in the side of the helium cryostat. The explosion of liquid helium that followed the rupture damaged several dozen magnets, shifting some of them by up to a foot and leaving soot in several kilometers of beam pipe. The quench protection system was overhauled in the following year, providing more robustness against such a failure and ensuring that, at the very least, if such an incident occurred again it would affect a much smaller number of magnets. Resistivity sensors throughout the ring were also upgraded to be an order of magnitude more sensitive.

In order to further protect the magnets against these catastrophic failures, an additional quench protection system will be implemented during a long shut down in 2011. "Jumper cables" will be attached to bridge all joints that could be missing solder, providing an additional route for the current if the joint ceases to be superconducting. These cables will further prevent damage to the magnet joint in the case of a quench, allowing for a faster recovery.

Before these additional measures are implemented, the LHC will be run at the reduced maximum beam energy of 7 TeV. Single beam operation was established on November 20, 2009, and collisions at a center-of-mass energy of 900 GeV were achieved on November 23. Collisions at a center-of-mass energy of 7 TeV were achieved on March 30, 2010. CERN hopes to deliver 1 fb^{-1} of integrated luminosity to the experiments during the 2010-2011 run before embarking on a year-long interruption to further enhance the quench protection system. In order to achieve this, the luminosities during the run must reach 10^{32}-10^{33} cm^{-2} s^{-1}, allowing >100 pb^{-1} per month towards the end of the run. For research presented in this degree, the amount of data delivered was 320±35 nb^{-1}.

Chapter 4 The ATLAS Detector

ATLAS (**A** **T**oroidal **L**HC **A**pparatu**S**) is one of two general-purpose detectors at the Large Hadron Collider [128]. Installation began in 1997 with the excavation of a cavern 92.5 m below ground at point 1 of the LHC. As shown in Figure 4.1, it is cylindrical in shape, 25 m in diameter, and 44 m long, and it weighs approximately 7000 tonnes. ATLAS is generally divided into three major components: the inner detector or "tracker," the calorimetry, and the muon system. The inner detector is emersed in a 2 T solenoidal field, and the muon system includes a toroidal magnet system that provides, on average, 2.5 T-m of bending power in the barrel and 5 T-m in the end-caps. There are also several forward detectors associated with ATLAS for measuring diffractive events, the luminosity, and the total cross section.

The ATLAS inner detector covers $|\eta| < 2.5$ and provides precision transverse momentum (p_T) measurements of charged tracks. The calorimetry is hermetic out to $|\eta| = 4.9$ and is $\sim 9 - 13$ interaction lengths thick, sufficient to capture 99% of the hadronic showers from single charged pions up to ∼500 GeV. The muon system covers the region $|\eta| < 2.7$ with different technologies to allow both fast triggering and precision position measurements. Each subdetector's position is known to a few mm thanks to laser alignment and surveying.

ATLAS includes a three-level trigger system for evaluating and recording only the most interesting events during a run. The trigger is configurable at every level to provide a constant stream of data under any beam conditions. The collected data are distributed around the world with the world-wide LHC Computing Grid [136] for reprocessing and analysis.

Most of ATLAS has been installed and integrated since 2008, and it has been recording cosmic-ray events and, more recently, proton-proton collision events at rates of up to 200 Hz. The millions of recorded events have proven invaluable for aligning and calibrating the detectors, exercising the data acquisition software, and commissioning the trigger. The data also have been used to begin validation of the detector simulation. As a result of the early data-taking campaigns, ATLAS has had good data quality and high data-taking efficiency for most of 2010.

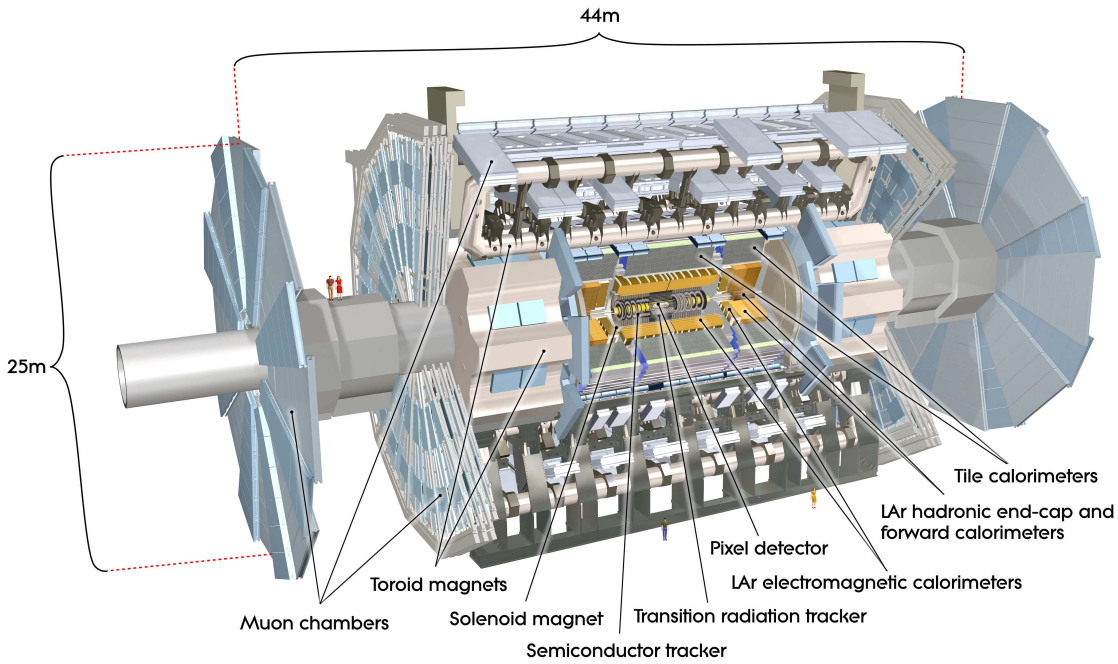


Figure 4.1: The main subsystems of ATLAS.

4.1 The Inner Detector

The ATLAS inner detector, shown in Figure 4.2, is composed of a silicon pixel tracker, a silicon strip tracker (SCT), and a transition radiation tracker (TRT). The entire inner detector sits inside a solenoid that provides a 2 T magnetic field. As charged particles traverse the active detector elements they leave a small deposits of energy, "hits," which are later recombined by the reconstruction software into tracks. The bending radius of the track provides a measurement of the transverse momentum (p_T in GeV) of the charged particle, $p_T = 0.3 \times B \times R$, where B is the magnetic field strength in Tesla and R is the bending radius in meters. The inner detector is used to determine a charged particle's transverse momentum, p_T, by recording several hits along its track (over 40 hits for most tracks) without affecting the particle's original momentum.

Attached to the pixel system is the beam conditions monitor (BCM). The BCM consists of eight diamond sensors 184 mm on either side of the interaction point. These sensors have very good time resolution (a signal rise time of 1 ns and average pulse length of 3 ns, compared to the LHC's typical bunch length of ~ 0.1 ns) and thus are able to distinguish particles arriving from the collision point from those arriving along the beam line. In the

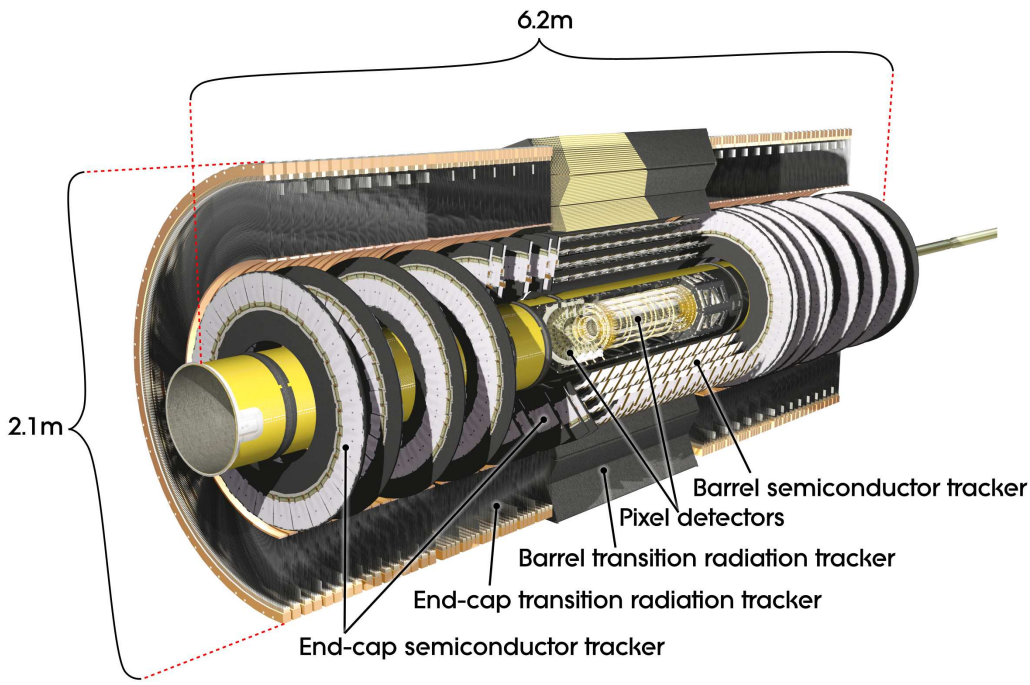


Figure 4.2: The ATLAS inner detector.

case of a beam incident, for example the beam scraping against the side of the beam pipe, a dangerously high flux of particles could be sent through ATLAS. To protect the detector, the BCM detects the high flux of particles not coming from the collision point and triggers the dump of the LHC beams. Once a beam dump signal has been sent, the beam is completely ejected from the main ring of the LHC in about three turns (∼250 μs). It is the only subdetector in ATLAS that is allowed to trigger a beam dump automatically. In 2009 and 2010, many beam dumps were triggered by the BCM, some as tests of the safety systems and some as a result of real incidents during beam commissioning.

The pixel detector is built directly onto the beryllium beam pipe in order to provide the best possible primary and secondary vertex resolution. It comprises three layers in the barrel at radii of 50.5, 88.5, and 122.5 mm, and three end-cap disks on either side of the center and positions $|z|$ = 495, 580, and 650 mm. The 250 μm thick silicon sensor "modules" are divided[1] into pixels 50 μm wide and 400 μm long, with 47232 pixels on each of the 1744 modules. With over 80 million pixels, the pixel detector makes up almost 90% of

[1]About 90% of the pixels have these "regular" dimensions. The remaining 10% are special pixels, which, owing to the electrical connections on the sensor, have irregular shapes and double or triple this typical area.

the total read out channels of ATLAS. A typical track with $\eta < 2.5$ will leave 3-4 hits in the pixel detector. Because of its proximity to the interaction region and the harsh radiation environment, the innermost layer of the pixel detector is expected to survive no more than five years at LHC design luminosity [137]. A program to construct a replacement for the inner layer is already underway [138].

The pixel sensors are essentially doped semiconductors with a high bias voltage. As a charged particle passes through the bulk, it creates electrons and holes which are collected at the surface of the silicon. A typical minimally ionizing particle passing through the sensor module will free 2×10^4 electrons. To reduce the number of hits coming from electronics noise, this analog signal is compared to a threshold, and only the time over threshold is recorded. Even in the case of a normally incident charged particle, several neighboring pixels may register a signal above threshold. These consecutive pixels are grouped into a single "cluster," which is treated as a single hit by the track reconstruction software. Clusters from noise typically have low time over threshold ($\sim$125 ns), whereas clusters from minimally ionizing charged particles are expected to have a time over threshold of $\sim$750 ns. The distribution of pixel cluster charge, after correction for the path length through the pixel sensor module, for clusters included in tracks with transverse momentum above 100 MeV, is shown in Figure 4.3. Within the geometric acceptance of the pixel detector, the hit efficiency is over 99%, but 2.5% of the modules in the detector were disabled during data-taking.

The SCT is similar to the pixel detector conceptually, but it uses 80 μm wide, 12 cm long silicon microstrips rather than pixels. The entire detector comprises 15912 sensor modules in four radial layers in the barrel, with radii of 299, 371, 443, and 514 mm, and six end-cap disks on either side of the interaction point at $|z|$ =890, 1091, 1350, 1771, 2115, and 2608 mm. The barrel layers, as well as the first, third, and sixth disks are all "double-layers." In these cases, two sensor modules are glued back-to-back, with the back sensor module offset at a small (40 mrad) angle. For detectors that are strip-like, rather than pixel-like, this double-layering is a common method to improve the spatial resolution in the long (i.e. poorly measured) dimension. The angle between the two layers is optimized for resolving "ghost" ambiguities when several charged particles traverse one double-layer [139]. The single-hit efficiency for the SCT is quite high, like that of the pixel detector, and only 0.97% of the sensor modules were disabled during data-taking.

The TRT, also called the "straw tube tracker," is composed of 4 mm diameter polyimide

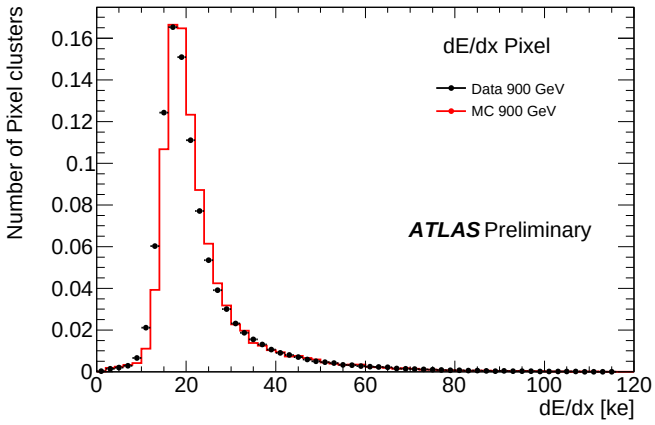


Figure 4.3: The charge freed (in thousands of electrons, ke) by a minimally ionizing particle traversing a pixel sensor module, after correction for the path length through the sensor. Tracks with transverse momentum above 100 MeV in 900 GeV collision data were used. For minimally ionizing particles, a peak at 20 ke is expected. Clusters from noise are expected to typically include less than 3 ke.

straws filled with a mixture of CO_2 and xenon. The straws are wrapped with a gold wire and act as the cathode of a drift tube, with a gold wire anode passing through the center of the straw. As charged particles pass, they ionize some of the gas. High voltage is applied across the straw, so that the charge is collected on the anode after some drift time. The drift time provides a precise measurement of the impact parameter of the track in the tube, but there is no resolution in the direction parallel to the wire (z in the barrel, R in the end-caps). Charged particles passing through the tubes also emit low-energy transition-radiation photons which are then absorbed by the gas. This transition radiation provides an independent discrimination between electrons and hadrons.

The coverage[2], spatial resolution, alignment tolerance, and expected number of hits on track of each of the subdetectors is shown in Table 4.1. The efficiency for finding a hit in silicon is near 99%. However, disabled portions of the detector and geometric acceptance reduce the expected number of hits. The standard metrics for inner detector alignment are the track residual errors (residuals). These residuals are calculated for each hit on a track by removing the hit, refitting the track through the remaining hits, and recording how far from the refit track the removed hit lies. The residuals in x for the pixel barrel and

[2] The innermost layer of the pixel barrel covers $\eta < 2.77$, and the outermost layer of the pixel end-caps covers $\eta < 2.69$. However, only tracks with $\eta <\sim 2.5$ include a sufficient number of pixel and SCT hits to be reconstructed with efficiency above ∼50%.

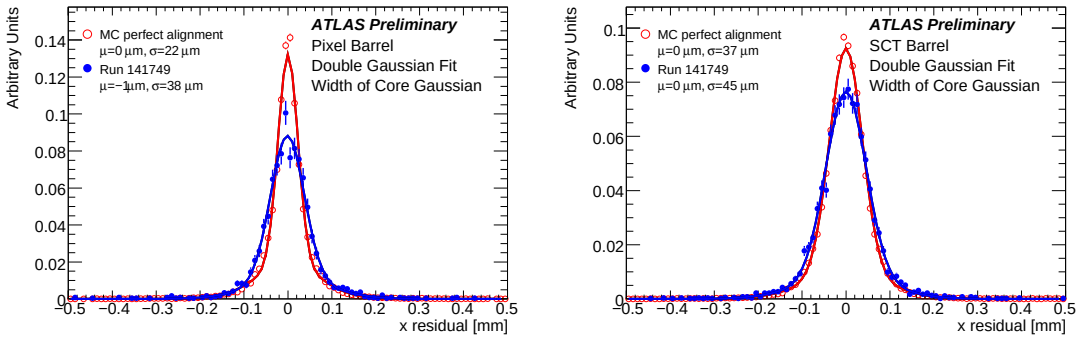


Figure 4.4: Track residual errors in the x direction calculated using 7 TeV data (blue) and Monte Carlo simulation (red) in the pixel tracker barrel (left) and SCT barrel (right).

SCT barrel calculated from tracks in a single run of 7 TeV collision data and Monte Carlo simulation are shown in Figure 4.4. In early 2010 (perfect Monte Carlo), ATLAS achieved track residual errors of 28 μm (23 μm) in the pixel tracker barrel and 43 μm (38 μm) in the SCT barrel.

Subdetector	**Coverage**	**Spatial resolution [μm]**	**Alignment tolerance [μm]**	**Hits per typical track**
Pixel Barrel	$\|\eta\| < 2.5$	10 (R-ϕ), 115 (z)	10 (R), 20 (z), 7 (R-ϕ)	3
Pixel End-caps	$2.0 < \|\eta\| < 2.5$	10 (R-ϕ), 115 (R)	20 (R), 100 (z), 7 (R-ϕ)	3
SCT Barrel	$\|\eta\| < 1.5$	17 (R-ϕ), 580 (z)	100 (R), 50 (z), 12 (R-ϕ)	8
SCT End-caps	$1.3 < \|\eta\| < 2.5$	17 (R-ϕ), 580 (R)	50 (R), 200 (z), 12 (R-ϕ)	8
TRT Barrel	$\|\eta\| < 1.0$	130	30 (R-ϕ)	35
TRT End-caps	$0.8 < \|\eta\| < 2.0$	130	30 (R-ϕ)	35

Table 4.1: Coverage, spatial resolution, alignment tolerance, and typical number of hits on track for each of the subdetectors in the ATLAS inner detector.

Both the pixels and SCT are cooled to ~ -10 C° by an evaporative cooling system using C_3F_8 as the coolant. The cooling is necessary to reduce the effect of radiation damage to the silicon. Because the TRT operates at room temperature, a set of insulators and heaters isolate the silicon detectors from the environment of ATLAS. In 2009 and 2010, several of the heaters were not functional. As a result, the outer layer of the SCT and inner straws of the TRT were operated at sub-optimal temperatures. In order to prevent condensation,

dry nitrogen is blown through the cooled portion of the inner detector. Because nitrogen would reduce the transition radiation signal in the TRT, the TRT is instead operated in a dry CO_2 environment.

For this analysis, jet shapes were only constructed using calorimetric objects. Therefore, the inner detector was only used for its primary vertex-finding ability. By requiring at least five tracks coming from a single point vertex near the luminous region in the inner detector, backgrounds from beam splash, beam gas, and cosmic ray events are highly suppressed. The vertex finding efficiency is almost 100% for events with more than two tracks, with a fake rate around 1% [140, 141]. The vertex resolution in the direction transverse to the beams is ~ 15 μm, and in the direction parallel to the beams it is ~ 50 μm, allowing good separation between multiple proton-proton interactions in a single event. Of course, these resolutions are highly dependent on the event topology. In the data considered here, $\sim$1% of the events had more than one proton-proton interaction. However, once the LHC reaches design luminosity, 23 proton-proton interactions are expected on average for every bunch crossing. The primary vertex is taken to be the one with the largest sum of the p_T^2 of all tracks originating from that vertex. Event selection using a vertex will be discussed further in Section 5.1.

4.2 Calorimetry

The ATLAS calorimetry comprises several different subsystems designed to cope with the significant variation in energy deposition over the entire range $|\eta| < 4.9$. The $\sim$20-40 radiation-length thick electromagnetic (EM) calorimeter is sufficient to contain electron and photon showers up to $\sim$1 TeV. It also absorbs almost 2/3 of a typical hadronic shower. The $\sim$8-12 interaction-length thick hadronic calorimetry contains most of the remainder of hadronic showers. The central hadronic calorimetry uses scintillating tiles (doped polystyrene) as the active medium with steel absorbers. Beyond the central region, liquid argon (LAr) is used as the active medium with lead, copper, or tungsten absorbers. The variation in absorber was dictated by density, cost, and construction difficulties. For example, tungsten, though the densest material, is very difficult to work with. In fact, the forward calorimeter of ATLAS is the first calorimeter to use tungsten as an absorber. An overview of the system is show in Figure 4.5.

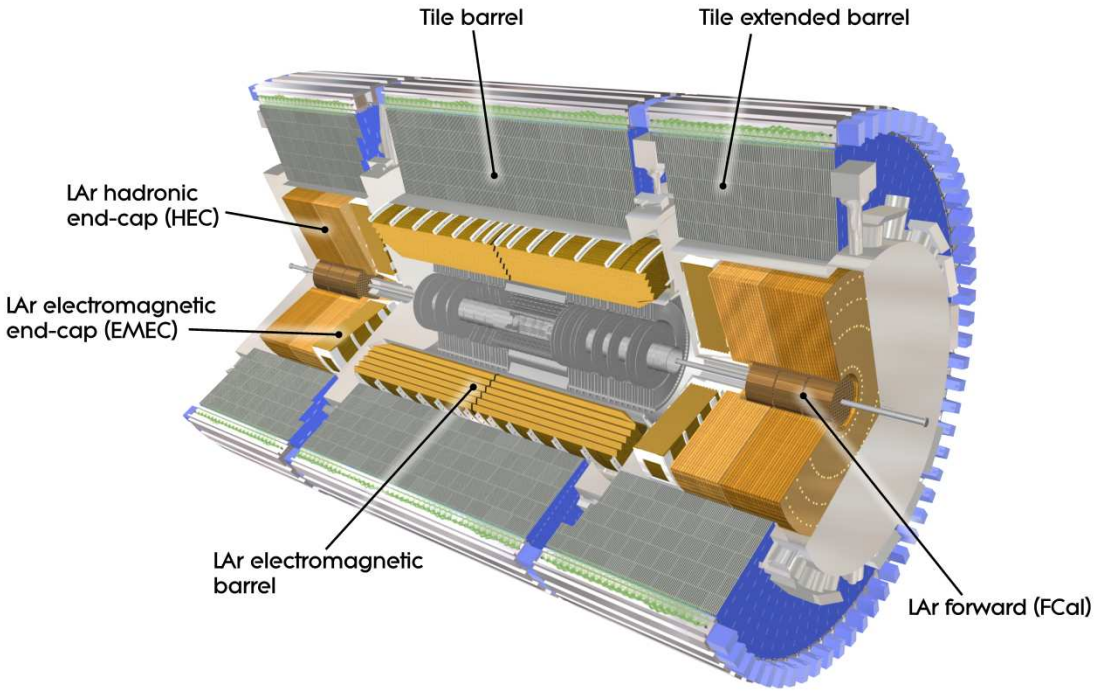


Figure 4.5: The calorimetry of ATLAS.

Both the LAr and the scintillating tile calorimeter technologies measure the energy of particle showers. Particles passing through the LAr calorimetry ionize the liquid argon. A high voltage is applied across the liquid argon, and the charge deposition is collected as a current in the cell. Because the LAr flows continuously, there is no reduction in signal strength due to radiation damage over the lifetime of the LHC. As they absorb radiation, scintillation light is produced in the tiles. The tiles' signal is transferred via optical fibers to photomultiplier tubes (PMTs) located in the support structure behind the active volume of the calorimeter. Due to radiation damage, the signal from the scintillating tiles is expected to decrease by several percent over the lifetime of the experiment. For calibration between runs, cesium sources have been installed which can scan across the detector to check its response (the ratio of the reconstructed signal to the "true" signal). Charge can be injected into a single cell to test and calibrate the read-out electronics, and lasers can provide light to test the optical connections and PMT response.

The EM barrel and end-cap calorimeters cover $|\eta| < 1.475$ and $1.375 < |\eta| < 3.2$ with LAr and lead samplings in an "accordion" geometry. The layers of lead wave in the ϕ

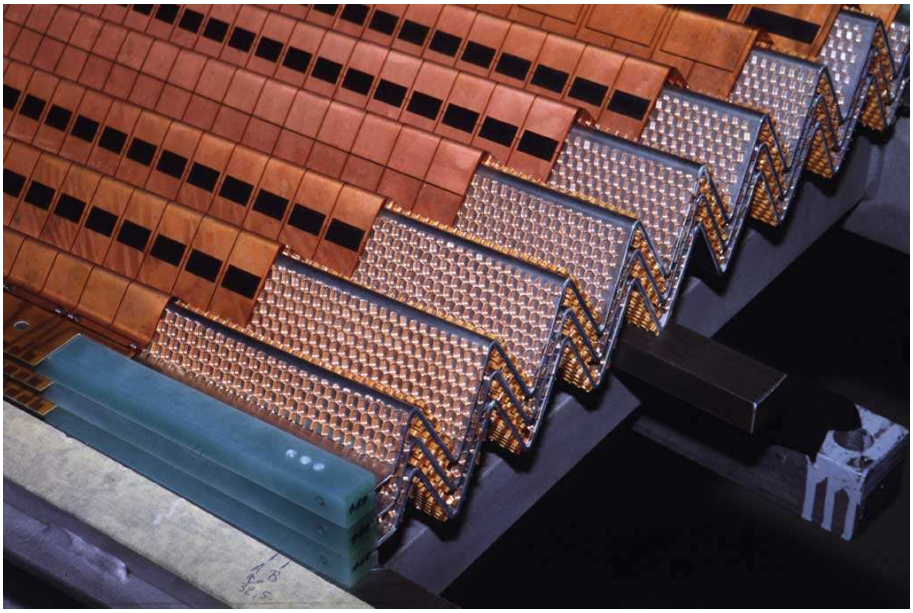


Figure 4.6: A portion of the ATLAS barrel electromagnetic calorimetry during assembly. The "accordion" geometry can be seen, where the center of the detector is to the bottom left-hand corner of the picture and the back of the detector (towards the hadronic calorimetry) is in the top right-hand corner. Sheets of lead absorber wave in the ϕ direction, providing full coverage in ϕ without any cracks.

direction, so that there are no cracks in the detector coverage in ϕ, and are supported with 0.2 mm thick stainless steel sheets for rigidity. A portion of the barrel detector during assembly is shown in Figure 4.6. In the barrel, the lead thickness changes from 1.53 mm below $|\eta| = 0.8$ to 1.13 mm above. In the end-cap, the thickness changes from 1.7 mm below $|\eta| = 2.5$ to 2.2 mm above. There are three longitudinal sampling layers in the EM calorimetry up to $|\eta| = 2.5$, two up to $|\eta| = 3.2$, and one in the forward calorimeter at large $|\eta|$. In the central region, the first layer is finely segmented in η and provides a good position measurement for the shower, the second collects the majority of an EM shower's signal, and the third contains the tail of the EM shower.

The central hadronic calorimeter includes a long barrel covering $|\eta| < 1.0$ and an extended barrels on each side covering $0.8 < |\eta| < 1.7$. It is made of 3 mm thick scintillating tiles in steel absorber, with 4-5 mm of steel between tiles. The tiles are oriented with the thin side in the z-direction. The calorimeter is divided into 64 wedges in ϕ, and the tiles are read out in groups to form three longitudinal samplings. A steel yoke around the outside, at $r \sim 4.25$ m, provides support for the barrel hadronic calorimeter as well as a return for the

solenoid field. As a result, there is a negligible magnetic field within the LAr calorimeter volume and a small field in the tile calorimeter volume.

The end-cap hadronic calorimeter covers $1.5 < |\eta| < 3.2$ with two longitudinal samplings in each of two wheels on both sides of the interaction point (i.e. four samplings per end-cap). The front (back) wheels are made from 32 ϕ-wedges with 24 (16) 25 mm (50 mm) thick copper plates surrounding 8.5 mm thick gaps filled with LAr. These gaps are divided into four drift regions that are read out with separate electrodes, providing some redundancy and allowing a shorter signal integration time.

Because most of the central calorimetry sits behind the cryostat, the solenoid, and the 1-4 radiation-lengths thick inner detector, EM showers begin to develop well before they are measured in the calorimeter. In order to measure and correct for these losses, up to $|\eta| = 1.8$ there is an additional presampler layer in front of the sampling portion (i.e. accordion) of the calorimetry. The presampler is 11 mm (5 mm) thick in the barrel (end-cap) and includes fine segmentation in η. Unlike the remainder of the calorimetry, the presampler has no absorber layer. In practice, it behaves almost like a single-layer LAr tracker.

In order to provide space for the routing of services from and access to the inner detector, the barrel and end-cap LAr calorimeters are housed in separate cryostats, with services running between them. To preserve energy measurements and correct for losses in the gap, scintillators were installed on the outside of the cryostats covering $1.0 < |\eta| < 1.2$ and $1.2 < |\eta| < 1.6$. A modified portion of a tile module is used for the same function between $0.8 < |\eta| < 1.0$ (between the tile barrel and extended barrel calorimeters) and simultaneously reduces the particle flux in the muon system by physically blocking part of the gap between the calorimeter pieces.

The forward calorimeter covers $3.1 < |\eta| < 4.9$ and consists of three 45 cm thick, independent modules, one EM and two hadronic. The geometry, as well as a schematic layout of the forward calorimeter, is shown in Figure 4.7. In the EM module, 12260 copper rods are inserted into holes drilled in a stack of copper plates. The 0.269 mm gap between the rods and the plates surrounding them is filled with LAr. In the hadronic modules, 10200 and 8224 tungsten rods are inserted into holes drilled in a stack of tungsten plates with copper plates on the ends. Gaps of 0.376 mm and 0.508 mm between the rods and the tubes surrounding them are filled with LAr. Each rod acts as an electrode, with several electrodes summed together at the detector in a single read-out channel.

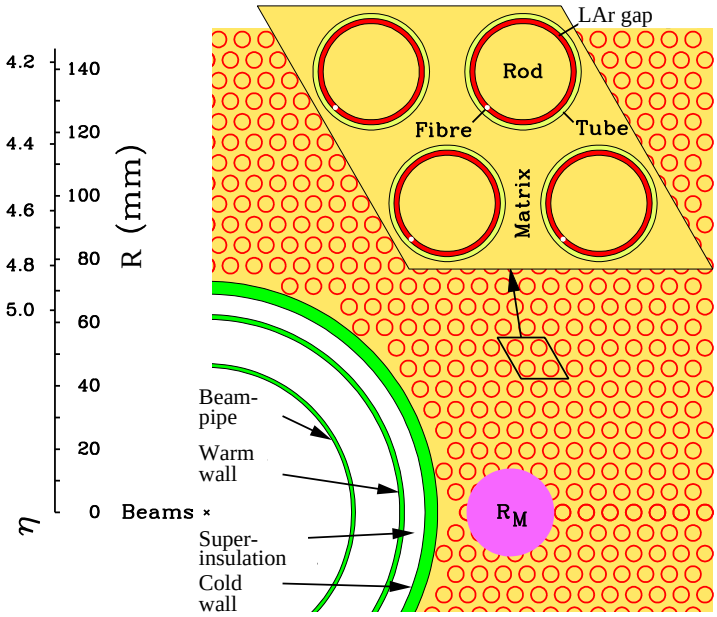


Figure 4.7: A schematic diagram of the ATLAS forward calorimeter (FCAL). The electromagnetic Molière radius (R_M) is shown in magenta.

The calorimeter is read out in cells that vary in size as a function of η and depth in the calorimeter. Cell sizes throughout the calorimeter are summarized in Table 4.2. For use in the hardware trigger and jet reconstruction, these cells are grouped into "towers" that always have an area of 0.1×0.1 in $\eta - \phi$ space. For cells larger than 0.1×0.1 in $\eta - \phi$ space, the cell's energy is evenly divided among the towers to which it contributes. These cells and towers are the smallest experimental objects from which jet shapes can be constructed. Previous experiments used calorimetric objects with granularity similar to the towers for the evaluation of jet shapes.

Two essentially separate read-out chains are maintained: one for the hardware trigger, and one for the remainder of the trigger and the offline reconstruction. The main read-out chain of the LAr calorimeter divides the calorimeter signal into 25 ns samples and passes each sample into a 12-bit analog-to-digital converter (ADC). Figure 4.8 shows example signals from the EM barrel calorimeter and the hadronic end-cap calorimeter, taken from a cosmic ray event, along with the predicted pulse shape [142]. At this stage, in order to maintain the full dynamic range of the calorimeter cell's energy read out, the signal is digitized using three gains (1, 9.3, and 93). The gain setting ultimately used to reconstruct the signal is chosen either at read-out time or later in the reconstruction process. Under

Subdetector	**Coverage**	**Cell Size**
EM Barrel	$\|\eta\| < 1.475$	0.025×0.1 in $\eta - \phi$ (Presampler)
		$0.025/8 \times 0.1$ in $\eta - \phi$ (Layer 1, $\|\eta\| < 1.4$)
		0.025×0.025 in $\eta - \phi$ (Layer 1, $1.4 < \|\eta\|$)
		0.025×0.025 in $\eta - \phi$ (Layer 2, $\|\eta\| < 1.4$)
		0.025×0.025 in $\eta - \phi$ (Layer 2, $1.4 < \|\eta\|$)
		0.050×0.025 in $\eta - \phi$ (Layer 3)
EM End-cap	$1.375 < \|\eta\| < 3.2$	0.25×0.1 in $\eta - \phi$ (Presampler)
		0.05×0.1 in $\eta - \phi$ (Layer 1, $\|\eta\| < 1.425$)
		0.025×0.1 in $\eta - \phi$ (Layer 1, $1.425 < \|\eta\| < 1.5$)
		$0.025/8 \times 0.1$ in $\eta - \phi$ (Layer 1, $1.5 < \|\eta\| < 1.8$)
		$0.025/6 \times 0.1$ in $\eta - \phi$ (Layer 1, $1,8 < \|\eta\| < 2.0$)
		$0.025/4 \times 0.1$ in $\eta - \phi$ (Layer 1, $2.0 < \|\eta\| < 2.4$)
		0.025×0.1 in $\eta - \phi$ (Layer 1, $2.4 < \|\eta\| < 2.5$)
		0.1×0.1 in $\eta - \phi$ (Layer 1, $2.5 < \|\eta\|$)
		0.050×0.025 in $\eta - \phi$ (Layer 2, $\|\eta\| < 1.425$)
		0.025×0.025 in $\eta - \phi$ (Layer 2, $1.425 < \|\eta\| < 2.5$)
		0.1×0.1 in $\eta - \phi$ (Layer 2, $2.5 < \|\eta\|$)
		0.050×0.025 in $\eta - \phi$ (Layer 3)
Tile Barrel	$\|\eta\| < 1.7$	0.1×0.1 in $\eta - \phi$ (Layer 1 and 2)
		0.1×0.1 in $\eta - \phi$ (Layer 3)
Hadronic	$1.5 < \|\eta\| < 3.2$	0.1×0.1 in $\eta - \phi$ ($\|\eta\| < 2.5$)
End-cap		0.2×0.1 in $\eta - \phi$ ($\|\eta\| > 2.5$)
Forward	$3.1 < \|\eta\| < 4.9$	3.0×2.5 cm^2 in $x - y$ (Layer 1, $3.15 < \|\eta\| < 4.30$)
Calorimeter		0.8×0.6 cm^2 in $x - y$ (Layer 1 elsewhere)
		3.3×4.2 cm^2 in $x - y$ (Layer 2, $3.24 < \|\eta\| < 4.50$)
		0.8×1.1 cm^2 in $x - y$ (Layer 2 elsewhere)
		5.4×4.7 cm^2 in $x - y$ (Layer 3, $3.32 < \|\eta\| < 4.60$)
		1.4×1.2 cm^2 in $x - y$ (Layer 3 elsewhere)

Table 4.2: Read-out cell sizes throughout the ATLAS calorimeter. In layer one of the EM calorimetry, cells comprise up to eight strips in η which are individually read out. Cells are grouped into towers for use by the trigger which are always 0.1×0.1 in $\eta - \phi$ space.

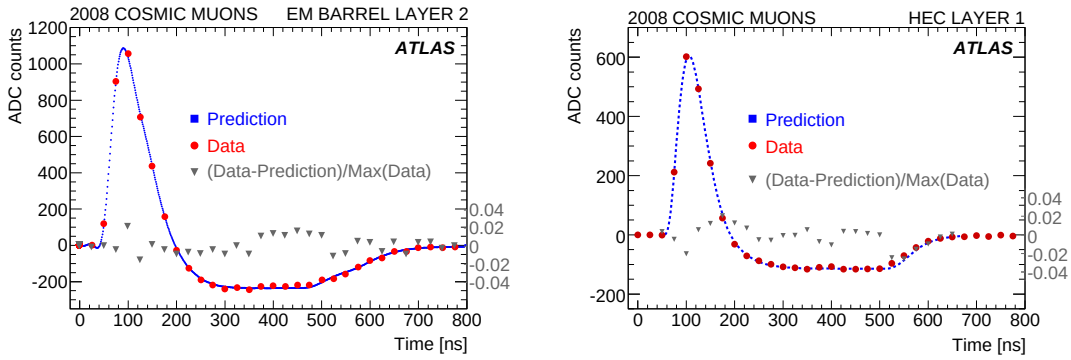


Figure 4.8: Examples of calorimeter signals in the EM barrel calorimeter (left) and the hadronic end-cap calorimeter (right), taken from a cosmic ray event, along with the predicted pulse shape. All 32 samples are shown as red points. The difference between the prediction and measured signal shape (grey) varies by less than 4%. During data taking at design luminosity, in order to reduce the amount of information that must be transferred off of the detector, only five samples will be read out. The pulse length is dominated by the drift time, the electronics' response, and the time required for the shower to develop. The amount of energy deposited in the cell affects the amplitude.

normal circumstances, five samples are read out, the first one including the rising edge of the showers, and the last required to catch a sufficient fraction of the showers' tails. For special calibration runs, however, as many as 32 samples may be read out, in order to evaluate the full pulse shape and electronics' recovery after a signal spike.

The calibrated energy deposited in the cell is derived from the current signal in that cell according to [143]:

$$E_{\text{cell}} = \frac{1}{f_{I/E}} \times F_{\text{gain}} \left(\sum_{\text{samples}} \text{OF}_{\text{sample,gain}} \left(S_{\text{sample}} - P_{\text{gain}} \right) \right) \tag{4.1}$$

A pedestal, P_{gain}, is subtracted from the signal in each cell, S_{sample}, and then the samples are re-summed using optimal filtering coefficients, $\text{OF}_{\text{sample,gain}}$, designed to suppress electronic noise as much as possible. As these filtering coefficients attempt to reconstruct the total pulse time and amplitude from a limited number of points, it is possible for them to reconstruct a signal with negative energy. Next, the sum (in ADC counts) is input to a function, F_{gain}, which converts the ADC counts back into a current. The relationship between ADC counts and current is non-linear, so F_{gain} is a fourth order polynomial with coefficients that are calculated individually for each calorimeter cell. This polynomial func-

tion results in ∼20 MeV residual uncertainties on the calibration of the cell energy [143]. Finally, a conversion factor, $f_{I/E}$, is applied to convert the signal into an energy.

This calibration is sufficient only to reconstruct the "visible" energy deposited in the cell. The reconstructed energy must then be scaled to account for the fraction of the detector in which signal is measured (i.e. the fraction of LAr or scintillating tile), called the "sampling fraction." The sampling fraction of the ATLAS calorimetry changes as a function of η because of the changing physical structure of the detector. The smaller sampling fraction in the more forward regions result in a signal that is approximately constant with p_T. After the adjustment for sampling fraction, the resulting energy is at the "EM scale." Once passive material (e.g. the material in the inner detector and the cryostats) have been accounted for, no further correction should be necessary to measure the energy of electrons and photons.

The ATLAS calorimetry is non-compensating, meaning that a hadron will deposit a smaller fraction of its energy in the active portion of the calorimeter, on average, than an electron or photon will. The difference must be taken into account by an additional correction applied to hadronic objects. There are several ways to select hadronic objects and correct their energy. This final calibration step will be described in Section 5.2.

One of the more significant calorimeter problems during 2009 and 2010 was the break down of many of the on-detector laser transmitters (OTXs) in the LAr calorimeter. The detailed energy deposition information is transmitted by 1524 lasers from the on-detector electronics to the computer racks away from the detector. A single OTX transmits the signal for cells in only one calorimeter layer, and the number of cells in the layer that it is responsible for varies considerably between OTXs. Several dozen of these lasers ceased to give light over the course of the data-taking period, and the energy deposition information was not available for the affected portions of the calorimeter. Because a single OTX is only responsible for the read-out of a single radial layer of the calorimeter, some are more critical than others. Loss of an OTX reading out in the third EM calorimeter layer is far less important than the loss of an OTX reading out in the second layer. The read-out of the calorimeter hardware trigger, however, is entirely analog. This provided an approximately redundant detector read-out, albeit with reduced granularity and worse noise than the optical read-out system. During reconstruction in the future, the analog trigger read-out information will be used as a replacement for any missing detector regions. The trigger towers have considerably more noise than the towers in the standard read-out. In order to

not dramatically affect the amount of noise in the calorimeter, only trigger towers with at least 2 GeV of transverse energy are used for this correction.

In the tile calorimeter, several low voltage power supplies and ϕ-wedge read-out systems failed during the 2009-2010 data taking run. Both these problems result in an entire ϕ-wedge (i.e. $\frac{1}{64}$ of the ϕ-coverage), or "drawer," from a single partition (i.e. half barrel or extended barrel on one side of the detector) being disabled, with the information from every sampling in that ϕ-wedge lost. The dead tile module scenarios were studied in some detail [144, 145]. Because the majority of a shower's energy deposition is caught by the LAr calorimeter, failures in the tile calorimeter do not have a dramatic effect on the measurement of a shower's energy, provided there are no failures in adjacent modules. These dead cells were corrected using the average of their neighbors' energies, as were stray dead cells throughout the calorimeter.

The performance of the various regions of the calorimeter were studied extensively using test beams at CERN from 2002 to 2006 [128, 146, 147, 148]. The linearity of the calorimeter was measured to be better than $\pm 0.1\%$ from 15 to 180 GeV, and the measured energy resolutions for single charged pions in the hadronic calorimeters were [128]:

$$\frac{\sigma(E)}{E} = \frac{0.564\ \text{GeV}^{1/2}}{\sqrt{E}} \oplus 0.055 \qquad \text{(Barrel)}$$

$$\frac{\sigma(E)}{E} = \frac{0.706\ \text{GeV}^{1/2}}{\sqrt{E}} \oplus 0.058 \qquad \text{(End-cap)}$$

$$\frac{\sigma(E)}{E} = \frac{0.942\ \text{GeV}^{1/2}}{\sqrt{E}} \oplus 0.075 \qquad \text{(Forward)}$$

All of these resolutions were found to be in agreement with the Monte Carlo simulation prediction for the energy resolution. However, these test beams used single particles to gauge the response and resolution of the calorimeter and were therefore not entirely indicative of the performance for complex objects like jets. By most metrics, the data collected through the middle of 2010 were consistent with the simulation results. Jet performance, calorimeter noise, and cleaning cuts will be discussed more extensively in Chapter 5.

4.3 Muon System

ATLAS uses four independent detector technologies for the muon system, shown in Figure 4.9. Muons frequently indicate an interesting event, and, therefore, a muon-based trigger is useful for selecting some new physics signals. However, in order to precisely measure the decays of new particles, it is necessary to make accurate measurements of each muon's momentum. Two technologies, the thin gap chambers (TGCs) and the resistive plate chambers (RPCs), are designed for rapid response. The other two, monitored drift tubes (MDTs) and cathode strip chambers (CSCs), are designed for precision measurements of track segments. The different technologies lead to considerably different operating points. For example, the TGC's require a high voltage of only 200 V, the MDTs and CSCs several kV, and the RPCs almost 10 kV. Because of the large physical separation of the muon chambers, each chamber provides more than one measurement of a muon track. These measurements help reject fake tracks and improve muon identification in regions with dead channels or detector inefficiencies.

The reconstruction efficiency and resolution of the muon spectrometer were measured using cosmic ray events in 2008 and 2009 [149]. The reconstruction efficiency, integrated over the detector acceptance, is ∼94%. At $\eta = 0$ there is a gap in the detector for cable routing. If the region of the detector near this track is excluded, the reconstruction efficiency is increased to 97%. The transverse momentum resolution was determined from this data to be

$$\frac{\sigma_{p_T}}{p_T} = \frac{0.29 \text{ GeV}}{p_T} \oplus 0.043 \oplus 4.1 \times 10^{-4} \text{ GeV}^{-1} \times p_T$$

for transverse momentum (p_T) between 5 and 400 GeV.

The TGCs contain two or three (depending on the type of chamber) 2.8 mm gaps filled with a mixture of CO_2 and C_5H_{12}. Wires run through the gap every 1.4 mm, collecting charge from ions created when muons pass through the gas. Otherwise, these chambers are similar in operational principles to the TRT straws in the inner detector. TGCs are physically arranged into end-cap disks and are used for triggering in the end-cap region ($1.05 < |\eta| < 2.4$).

The RPCs comprise a 2 mm gas gap between an anode plate and a cathode plate, with metallic strips for read-out on the outside of the plates. Two sets of read-out strips are

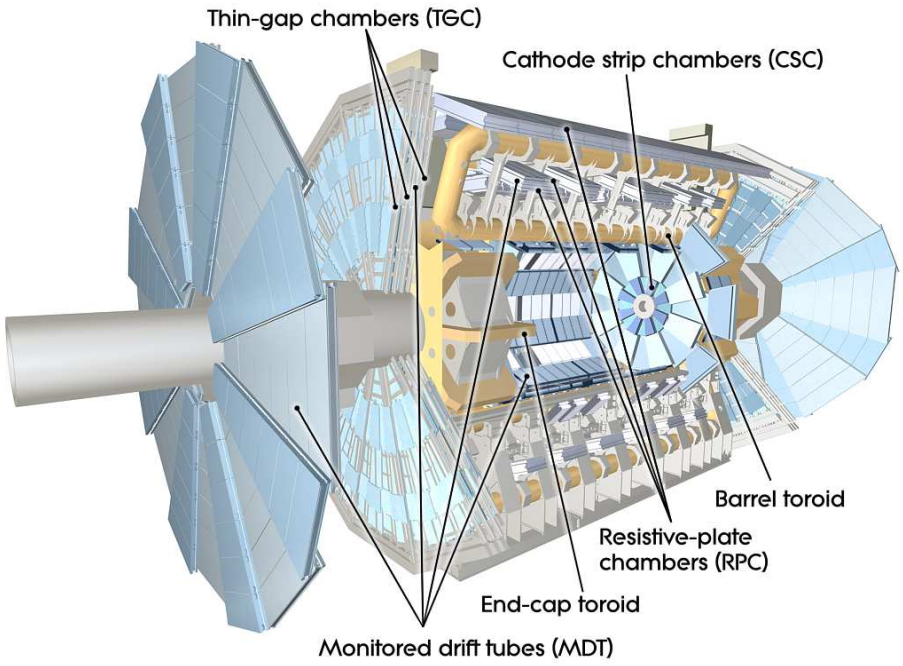


Figure 4.9: The muon system of ATLAS.

used, one in the ϕ-direction and one in the η-direction, to measure the coordinates of a hit. RPCs are arranged around the barrel in three double-layers, similar to those of the SCT, and are used for triggering in the central region ($|\eta| < 1.05$). The first two layers are close together (7820 and 8365 mm in R), providing a trigger for muons with p_T from ∼6 to 9 GeV. The third layer, considerably further out in the detector (10229 mm in R), is used only in the triggers for muons with p_T from $\sim$ 9 to 35 GeV. These triggers are over 99% efficient for muons in the geometric acceptance of the detectors.

The MDTs have three or four layers of 30 mm radius aluminum drift tubes on each side of the chambers. The drift tubes are filled with argon, CO_2, and water vapor and, like the TRT straws in the inner detector, have an anode wire running down the middle of them. The chambers vary in length from 1 to 6 m and in width from 1 to 2 m, depending on their position in the detector. The MDTs cover the region $|\eta| < 2.4$. Because of the physical size of the chambers, deformations due to gravitational sagging or thermal swelling can be significant. Therefore, a laser alignment system is built-in to help achieve the desired 80 μm position resolution per tube layer.

The CSCs cover the high rate region $2.0 < |\eta| < 2.7$ with one end-cap disk on each side

of the detector containing sixteen chambers. The chambers comprise four layers of argon and CO_2 gas with anode wires running radially. On one side of the wires a set of cathode strips runs radially, and on the other side a set of cathode strips runs transversely. Unlike most of the other detectors, the anode wires are not read out. Only the cathode strips are read out, with one set providing an η coordinate measurement and the other set a ϕ coordinate measurement for each hit.

The muon system is built around three air-core toroidal magnets, also shown in Figure 4.9, that bend the muon track so that its momentum can be measured. The two end-cap toroids fit in either end of the barrel toroid. The magnets provide, on average, 2.5 T-m of bending power in the central ($|\eta| < 1.4$) region and 5 T-m in the end-cap ($1.6 < |\eta| < 2.7$) region. Because the toroidal magnet system of the muon system is completely independent of the solenoid in the inner detector, ATLAS is able to acquire two independent measurements of a muon's momentum as it passes out of the detector.

Due to its size and the environment of the LHC, the muon system needs to cope with a significant amount of background radiation [150]. This radiation, mostly consisting of low-energy photons and neutrons, creates fake noise hits in the muon chambers beyond the background electronics noise. Because the radiation can come from beam gas interactions, beam halo interactions, very long-lived neutrons, and even activated materials in the ATLAS cavern, it is difficult to model and must, therefore, be determined from data. A program for acquiring such data was underway in 2009 and 2010, but it was not sufficiently advanced to be used for this analysis.

4.4 Forward Detectors

ATLAS includes three forward detectors for measuring luminosity and diffractive physics. LUCID (LUminosity measurement using Cherenkov Integrating Detector) is a relative luminosity detector that resides in the forward shielding near the muon end-caps, 17 m from the interaction point. The ALFA (Absolute Luminosity For ATLAS) detector comprises a scintillating fiber tracker inside Roman pots 240 m from the interaction point. The ZDC (Zero Degree Calorimeter) measures neutral particles with quartz rods inside of tungsten plates 140 m from the interaction point. With these separate systems, as well as complementary measurements using the forward calorimeter, ATLAS achieved an absolute luminosity

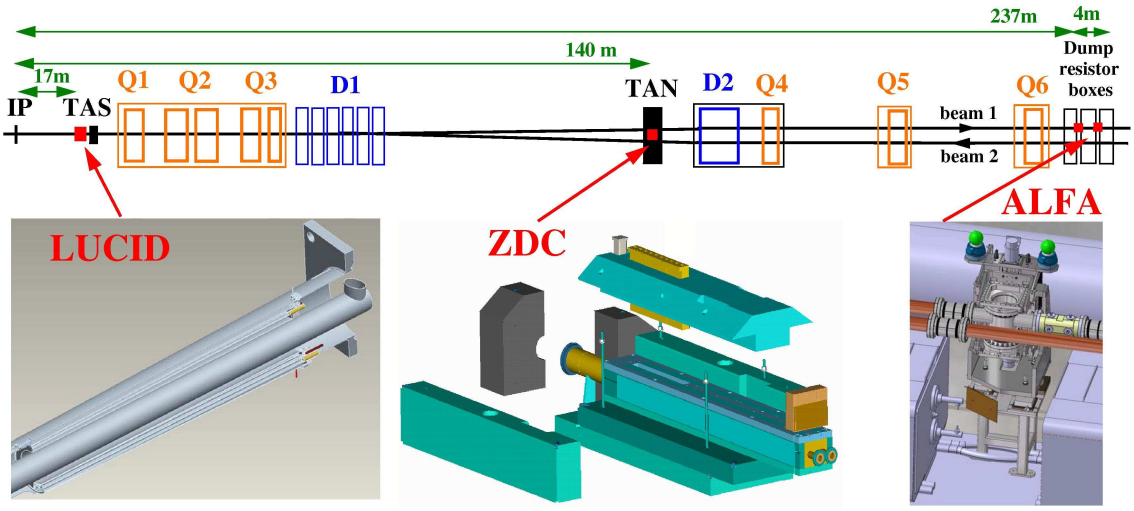


Figure 4.10: The forward detectors of ATLAS. From left to right, and closest to the interaction point to furthest, they are LUCID (LUminosity measurement using Cherenkov Integrating Detector), the ZDC (Zero Degree Calorimeter), and ALFA (Absolute Luminosity For ATLAS). Also shown in the picture are the various quadrupole (Q) and dipole (D) magnets in the beam line between the interaction point and ALFA. The Target Absorber for Neutrals (TAN) is the point at which the beams split into separate beam pipes.

measurement with an ∼11% systematic uncertainty in 2010 [151]. The layout of the three forward detectors with respect to the interaction point is show in Figure 4.10.

LUCID consists of forty (twenty on each side of ATLAS) 1.5 m long, 15 mm diameter aluminum tubes filled with C_4F_{10} at 1.2 to 1.4 bar. As a charged particle passes through the gas, it emits Cherenkov radiation, which travels down the tube and is collected by a photomultiplier tube (PMT). The light collected is proportional to the number of charged particles in the tube and so can provide a relative luminosity measurement and a collision activity trigger. Because of the high radiation in the detector at $|\eta| \sim 5.8$, the signal from the PMTs is expected to degrade over time. For high luminosity running (above 10^{33} cm^{-2} s^{-1}), the tubes will instead be read out via a bundle of quartz fibers leading to a PMT in a better radiation-shielded area.

In order to measure elastic scattering at very small angles, ALFA sits in Roman pots that can position the detectors as close as ∼1 mm away from the beam. ALFA comprises two Roman pots separated by four meters on both sides of the interaction point. Each Roman

pot houses ten double-sided fiber tracker modules with sixty-four 0.5 mm wide fibers. The detector measures elastic scattering and determines absolute luminosity during a run using the total scattering cross section. In order to make these measurements, ALFA requires better-than-normal beam divergence, and, therefore, can only be used during dedicated LHC runs. It will be used to calibrate LUCID.

The ZDC is designed to measure the centrality of heavy-ion collisions. It sits at the end of the straight section of beam pipe, just before the two proton beams split into separate beam lines, 140 m from the interaction point at $|\eta| \sim 8.3$. On each side of the ATLAS detector are one electromagnetic module and three hadronic modules. Each module contains layers of 1.5 mm thick quartz strips, read out by PMTs, sandwiched between layers of tungsten and steel. The thickness of the tungsten and steel layers and number of strips read out in a single channel distinguish the electromagnetic and hadronic modules. The ZDC is only used in proton-proton collisions with luminosities below 10^{33} cm^2 s^{-1} to avoid radiation damage to the quartz.

4.5 Data Acquisition System and Trigger

At even moderate luminosities in the LHC (one twentieth of design luminosity), protons will collide in ATLAS every 25 ns. Some 100 million channels in the ATLAS detector must be read out by the data acquisition software during LHC operation, resulting in ∼1.5 MB events. Without any filtering, ATLAS would need to process and record ∼60 terabytes of data every second, currently an impossible task. Because only a small fraction of the events can be recorded, these events must be quickly searched for interesting signatures. This presents a significant challenge to the ATLAS trigger and data acquisition system. A rapid decision must be made for each event. But rejected events are, of course, lost forever.

The rate at which ATLAS can record events is actually limited by its ability to transfer the events away from point 1. A steady transfer can be maintained at ∼700 MB/s from point 1 to the data processing center at CERN (Tier0), implying an ideal output rate of ∼400 Hz. At Tier0, there is sufficient computing power to process 200 Hz of data with the LHC providing collisions 70% of the time. Thus, some ∼200,000 events must be discarded for every one that is kept as design luminosity. Using on-site buffering capability, ATLAS can write at the nominal rate of 200 Hz for up to 24 hours without any connection to the

Tier0. This condition results in strict limits on the event size. If ATLAS were to record a float (eight bytes) for every channel in the detector, each event would require ∼800 MB of disk space. In order to keep the number of read-out channels to a minimum, most detectors apply zero-suppression, only reading out channels with a signal above some threshold. In the case of the pixel detector, the need is obvious: the expected number of hits per event is six orders of magnitude lower than the number of channels in the detector. This zero-suppression can be disabled for special runs, for example to understand data acquisition problems or monitor detector noise below threshold.

ATLAS has implemented a three-tier trigger system to handle the high-rate environment. At each level, physics objects are reconstructed with improved granularity and precision and over a larger fraction of the detector, culminating in complete event reconstruction in the final trigger stage. The level-one (L1) trigger is a configurable, pure-hardware trigger that is designed to make a decision on each event in under 2.5 μs and provide output at a rate up to 75 kHz. No inner detector information is used at this stage, and only very simple calorimetric and muon reconstruction information is available. The level-two (L2) triggers make a decision in under 40 ms and provide output at rates up to 3.5 kHz. The L2 triggers run a simplified version of the event reconstruction in the regions of interest defined by the calorimeter and muon systems. Improved selection criteria, such as distinguishing electrons from photons by track matching, and improved calibrations are applied. Some limited identification of b-jets is done, and the missing transverse energy measurement is refined, for example to include muons. In the level-three trigger, called the "event filter" (EF), the complete offline event reconstruction makes a decision in under 4 s and provides output at 200-400 Hz. The L2 and EF are both software triggers, unlike the L1 trigger, and they are together referred to as the "high-level trigger" (HLT). In 2009 and 2010, rates between 100 and 200 Hz were not uncommon.

A "menu" of possible trigger items is prepared for each data taking run. The menu defines a complete list of what trigger items will be evaluated, what values the parameters of those items will take, and how the lower-level trigger items map into higher-level triggers. One L1 item may seed many HLT triggers, and many L1 items may seed a single HLT trigger. A full sequence of triggers from L1 through the EF is called a trigger "chain." Some items are run unprescaled, meaning that any time an event is accepted by the trigger it will be passed on to the next level (or written out in the case of the EF). Others, in

particular low-p_T triggers, may be run with relatively high prescales, so that only some of the events which pass the trigger are accepted. An item may also be disabled by applying a negative prescale. In this way, a single menu can provide maximum flexibility and may be applicable for a large range of machine luminosities. Any of the HLT triggers can be run in pass-through mode, in which no events are rejected but each event is evaluated nonetheless. This feature is particularly useful when commissioning new trigger items. Because of the strict timing demands, if an event cannot be evaluated in the allotted time for each trigger stage, it is passed and flagged for later examination.

Some of the L1 triggers, for example some of those based on LUCID and the ZDC information, provide a 'yes' or 'no' for each event, which can then be prescaled to meet the desired rate. The calorimeter and muon systems instead define Regions of Interest (RoIs) in which interesting physics objects (electrons and photons, jets, muons, and τ-leptons) are contained. These RoIs are defined with an approximate p_T measurement for the object of interest. Their size is configurable (in the calorimeter, typically, they are cones of radius 0.7 in $\eta - \phi$), and, in the case of the calorimeter, their direction corresponds to the local maximum of energy deposition. The calorimeter also provides an uncalibrated total transverse energy and missing transverse energy at L1, uncorrected for muons in the event. Both simple triggers based on one object at a single p_T threshold and complex triggers based on several objects at various thresholds can be configured. The event information is retained in on-detector buffers until a decision is made. If the event passes one of the L1 triggers, the information is sent on to a dedicated computing farm where the HLT is evaluated.

Several simple L1 triggers exist for debugging, calibration, or monitoring purposes. Beam pickups, measuring the mirror current running along the inside of the beam pipe, on either side of the cavern provide a filled bucket trigger. With $1.5{\times}10^{11}$ protons in a single bunch, these pickups measure a $\sim$30 V signal. This trigger can be used for noise studies and trigger efficiency evaluation, to record background events without any trigger bias, or to record single beam events, for example. It can also be used to select only events that included two filled buckets. In 2009 and 2010, the beam pickup trigger was required in addition to most of the physics triggers, reducing the rate of beam halo, beam gas, and cosmic-ray events passing the physics triggers by at least three orders of magnitude. An additional random trigger can be used to select non-filled buckets. Either of these triggers

can be run at a low rate (e.g. $\sim$1 Hz) at L1 and in pass-through mode in the HLT, or they can be run at a high rate at L1 with a significant prescale at the HLT in order to test high-rate data-taking issues.

After the EF, the events are divided into "streams," each containing the output from several different trigger chains. A single stream may include, for example, all events passing any photon or electron trigger. All of the events passing the L1 calorimeter triggers (including jet, electron, photon, and τ-lepton triggers) were put into the "L1Calo" stream in the first half of 2010. Events may be added to several streams if they pass several different triggers. The number and contents of the streams are decided at the beginning of each run, but typically a "debugging" stream is reserved for events that could not be evaluated in the time allotted to a trigger. Additionally, streams called the "express stream" and "calibration stream" contain an assortment of events which are deemed interesting or useful for calibration of the subdetectors. In practice, $\sim$ 10 streams are included at any given time.

Eventually, these streams are transferred to the Tier0, where full offline event reconstruction is run, and the output is saved for distribution to computing centers around the world. During this re-reconstruction, additional algorithms are run, and additional data monitoring is done, which would not have satisfied the strict time requirements for online reconstruction. Additional output files may also be produced at this stage, including a first stage of additional processing and event filtering for some physics analysis groups. The express and calibration streams are processed first in order to provide new calibrations to the detectors within 24-hour periods. Any subsequent data reprocessing is done at various worldwide computing centers in order to unburden the Tier0.

A single data taking run in ATLAS can last for many hours. Typically, one long run is taken during an LHC fill, and, if necessary, between fills, the run is stopped for detector calibrations. These runs are divided into "luminosity blocks" that are a few minutes long each. Luminosity blocks are the smallest units of data for an analysis in ATLAS. Each luminosity block can be included or excluded in the final analysis.

Some changes to the run configuration can be made on the fly. For example, trigger prescales can be changed. Most detectors also have the ability to remove sections without stopping the data taking, in case a single section becomes problematic for either the trigger or the data read-out chain. After any of these transitions, a luminosity block boundary is introduced in order to provide a clear division between data taken prior to, during, and

after the transition. Other modifications, such as changes in detector configuration or the inclusion of additional subdetectors in data taking, can only be made between runs. At the beginning of an LHC fill, for safety reasons, the pixel detector operates with both module high voltage and signal preamplification off, and the SCT operates at a lowered high voltage. These conditions are changed during the run, after the beams have been declared stable, and background levels have been checked.

One of the most critical challenges to the data acquisition system in ATLAS is ensuring that all triggers and detectors are in sync with one another. Most detectors run on a 40 MHz clock, and the phases of all the systems must be adjusted so that a given clock cycle in all the subsystems corresponds to the same bunch crossing. Without this timing, subdetectors may associate a signal to the incorrect collision, or miss a collision's signal all together. The phase and timing is so precise that, as the beams are ramped from 450 GeV to 3.5 TeV, the clocks are actually sped up to keep sync with the crossings of the beam. The process of aligning the clocks of the various systems is called "timing in."

Prior to 2010, not all detectors were timed-in well. Therefore, it was necessary to read out several more bunch crossings' information around a trigger during early 2010 runs. This ensured that no data would be lost because signals were associated with the incorrect bunch crossing. The LAr calorimeters, for example, provided seven bunch crossings' data, and the pixel detector provided five bunch crossings' data. Once sufficiently well timed in, the LAr calorimeter will read out five bunch crossings, and the pixel detector will read out only one, except in the case of detector calibration. Because of the larger data volume of this extended read-out, it is necessary to introduce some "dead time" between successive L1 triggers. Dead time varied significantly over the 2009 and 2010 runs but was typically several hundred nanoseconds.

During the 2009 and 2010 runs, much of the trigger system needed to be commissioned and timed-in. Only L1 triggers were used for early running, and the HLT was run in pass-through mode. The beam pickup trigger was used first to ensure that the subdetectors were all writing out signals from the correct bunch crossing. Next, the simplest luminosity monitor triggers were included and timed-in. Finally, the L1 calorimeter and muon triggers were included. At each stage, the triggers had to be tuned to identify the correct bunch crossing. Only after these L1 triggers were timed in, HLT triggers were included and commissioned. Within 30 minutes of the start of the 7 TeV collisions in 2010, the HLT

was included and active online, though it was not rejecting physics events. Most triggers for interesting physics events, including all triggers used in this analysis, did not need to be prescaled. Therefore, the HLT was only used in pass-through mode for these triggers.

4.6 Offline Software and Data Processing

ATLAS's data set is far too large to be contained and processed entirely at the Tier0. Instead, the worldwide Large Computing Grid (LCG) [136] is used to distribute data around the world for processing and user access. Several copies can be maintained, so that if any single processing site is unavailable, the data can still be accessed.

In order to correctly reconstruct the data, a great deal of "conditions" information is necessary. ATLAS records this data constantly and distributes the necessary parts on the LCG. For example, the response of the LAr varies with temperature by 2% K^{-1}. It is, therefore, necessary to have accurate information on the temperature of the LAr when the data were recorded and to use that information for signal reconstruction. Similarly, the prescales of all trigger items and any trigger dead time are distributed as conditions information.

In order to understand detector effects in the data, cope with varying conditions, and test analyses of new physics, ATLAS has implemented a detailed detector simulation [152] based on the GEANT4 toolkit [153, 154]. The detector simulation has been tuned to agree with test beam data from both standalone system test beams and a "combined" test beam with a complete wedge of the detector [146, 147, 148, 155, 156]. Simple validation of the simulation is also done by, for example, weighing portions of the detector prior to installation in the pit and comparing the total weight measured to the sum of the masses of all the simulated components. By this measure, agreement within a few kg (better than 1%) in the inner detector but is off by several tons in the muon spectrometer. The simulated luminosity of the machine may be changed to match beam conditions for different runs, and triggers can be simulated.

Although it is possible to vary the simulated detector conditions according to the time variation of the conditions of the detector during data taking, for simplicity, a representative set of conditions is chosen and simulated. For slowly varying conditions (e.g. the number of dark optical transmitters in the calorimetry), some effort is made to have the exact data-

taking conditions present in the simulation. For more rapidly varying conditions (e.g. the temperature of the muon systems), only some average set of conditions are included. For example, the position of the beam spot may change by several mm between fills. Because it is important to evaluate the detector's performance for not only an ideal beam spot at the geometric center of ATLAS, but also for beam spots that are displaced by several mm, two representative beam spots were simulated. Of course, neither of these simulated beam spots may correspond to the measured beam spot during a particular run. The pixel and SCT modules that were disabled during the data taking period were also disabled in the simulation, but dead pixels, for example, are distributed randomly throughout the remaining modules (i.e. no attempt is made to mask exactly those channels which are dead on the detector, since the set of masked channels may change frequently). Similarly, the LAr optical transmitters (OTXs) and tile drawers that were disabled at the beginning of the data taking period are also disabled in the simulation. Thus, for example, the OTXs disabled in the simulation are identical to those disabled during the first run of the data taking period. However, because additional OTXs and tile drawers were disabled throughout the data taking period, the average number disabled in the data differs from the average number disabled in the simulation. The number of OTXs and tile drawers disabled in the simulation and in maximum number disabled in the data are shown in Table 4.3, along with their approximate extent in $\eta - \phi$. As the OTXs only read out a single layer of the LAr calorimeter, the extent of dead OTXs listed here is pessimistic.

Subdetector	**Disabled in Simulation**	**Detector Coverage**	**Disabled for Data Taking**	**Detector Coverage**
Tile long barrel drawers	5	∼7.8%	5	∼7.8%
Tile extended barrel drawers	2	∼3.1%	2	∼3.1%
LAr EM Barrel OTXs	8	∼2.2%	14	∼7.8%
LAr EM End-cap OTXs	8	∼3.3%	12	∼5.2%

Table 4.3: The number of LAr calorimeter optical transmitters (OTXs) and tile calorimeter drawers disabled in the simulation and the maximum number in the data used for this analysis, along with their detector coverage. The majority of the discrepancy between data and simulation in the LAr EM barrel comes from a single OTX presampler covering a 1.5×0.2 region in $\eta - \phi$.

Calculating the magnetic field of ATLAS is a complicated task, particularly because of distortions from the iron and steel around the detector. Therefore, a field map is created for

use in both the simulation and the track reconstruction with variable meshing to follow the measured field as closely as possible. This map is matched to measurements from several thousand hall cubes throughout the muon system and to the detailed survey of the solenoid field taken before the inner detector was installed. The solenoidal and toroid magnets' currents, however, may vary by a small amount from the nominal current during a run. To adjust for this effect, the magnetic field map is automatically rescaled according to the current measured at the appropriate time. Changes in magnet currents larger than about 5 A require a freshly calculated map.

4.7 Conditions and Data Quality for the 2010 Physics Runs

In order to provide useful information about the quality of the detector output during each physics run, ATLAS has implemented a set of data quality flags [157]. These flags take one of several values: green (good), yellow, red (bad), black (detector off during the run), and gray (undefined). Generally, it is left to the detector group to define the precise meaning of the flags. For the calorimetry, in general, a red flag means that something changed during the luminosity block (e.g. a portion of the detector was disabled for part of that luminosity block) or that there was some unrecoverable data acquisition problem. A green flag means that problems with the detector were understood. A yellow or gray flag means that the quality of the data is uncertain, and a detector expert needs to check the run. A yellow flag may also indicate that the data needs to be reprocessed, for example to mask noisy channels.

The number of read-out channels for each subdetector and the fraction of the detector providing useful data during 2009-2010 is given in Table 4.4 [158]. During stable beam running, detector up time was 98.2%. Table 4.4 also shows the fraction of the data taken with each detector that was considered "good" after offline reprocessing. The integrated luminosity as a function of day of the 2010 run is shown in Figure 4.11, separately for the delivered integrated luminosity, and recorded integrated luminosity with all subdetectors enabled. The difference between the latter two is approximately the time required to turn on all the detectors from "safe" mode. The total recorded luminosity for this data sample is 312±34 nb^{-1} [151].

Subdetector	**Read-out channels**	**Operational Percentage**	**Percentage of "Good" Data**
Silicon pixel tracker	80×10^6	97.5%	97.3%
SCT	6.3×10^6	99.0%	98.4%
TRT	350×10^3	98.0%	100%
LAr EM calorimeter	170×10^3	98.5%	94.5%
Hadronic end-cap calorimeter	5600	99.9%	99.2%
Forward LAr calorimeter	3500	100%	99.2%
Tile calorimeter	9800	97.3%	100%
TGC muon chambers	320×10^3	98.8%	98.5%
RPC muon chambers	370×10^3	97.3%	97.1%
MDT muon chambers	350×10^3	99.7%	98.4%
CSC muon chambers	31×10^3	98.5%	97.9%

Table 4.4: The number of read-out channels, percentage of good channels, percentage of up time during collisions with stable beams, and percentage of good data in each subdetector during the 2010 physics runs.

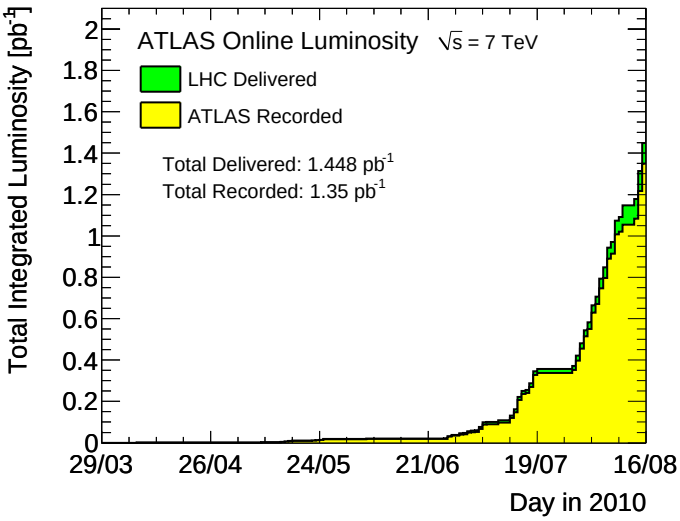


Figure 4.11: The integrated luminosity recorded by ATLAS over the data-taking period included in this analysis. Shown separately are the delivered integrated luminosity (green) and the recorded integrated luminosity with stable LHC beams (yellow). All luminosity estimates have a systematic error of 11%.

Chapter 5 Analysis

Data included in this analysis were collected from April to July 2010 (ATLAS data taking periods "A" through "D") and comprised 312±34 nb^{-1} [151]. The detector status during this period is described in detail in Chapter 4 and Ref. [159]. Typically, 95% of the data recorded by ATLAS were usable for physics analysis. The level one calorimeter trigger was not well timed-in for first few runs of 7 TeV collisions in 2010 [160]. As they comprise only ∼0.5% of the data sample, these runs were excluded from the analysis.

ATLAS has chosen to focus studies on Anti-k_t jets in early data [85]. Therefore, for this analysis, Anti-k_t jets with D=0.6 were used. Jets were reconstructed using the FASTJET [161] 2.4.1p1 implementation of the Anti-k_t jet algorithm. "Hadron level" (also called "generator-level") jets and jet shapes were built from stable hadrons produced by the generator, prior to the interference of any detector effects. Unless explicitly stated otherwise, experimental jets and jet shapes were built from calorimeter towers with topological noise suppression, using the uncalibrated (EM scale) energy of the tower. The jet energy is calibrated using a p_T-and η-dependent scale factor (see Section 5.2).

All data and Monte Carlo simulation used were a part of the ATLAS "May 2010" reprocessing campaign. The analysis was done using ATLAS's ATHENA [162] software version AtlasProduction-15.6.9.10. Generation used PYTHIA [54] version 421.2 with the "new" (p_T-ordered) shower and the ATLAS tune MC09 [63]. For comparisons, HERWIG [56, 57] version 6.510.3 with JIMMY [163] 4.31.3, ALPGEN [42] version 2.13 with HERWIG / JIMMY, and HERWIG++ [58] version 2.4.2 were used. PYTHIA used the string fragmentation model for hadronization, and HERWIG and HERWIG++ included the cluster model for hadronization. For the PYTHIA, HERWIG, and HERWIG++ generators, MRST modified LO PDFs (see Chapter 2) were used [37] from the LHAPDF library [39]. These are leading order PDFs, with leading order α_s, modified to approximate next-to-leading order results in many observables. ALPGEN used the CTEQ6L1 PDFs and the "old" ATLAS MC08 tune for HERWIG / JIMMY. Neither the ATLAS MC09 tunes nor the HERWIG++ tune used included jet shape measurements from the Tevatron as a part of their input data set. The full detector simulation was done using GEANT4 [153, 154] version 9.2p2.

Because of the rapidly falling jet transverse momentum distribution, Monte Carlo generation is done in several disjoint regions of phase space. Events were generated in bins of the transverse momentum of the leading parton coming from the hard scatter. A similar number of events were generated in each bin, providing a more uniform distribution of events in the jet p_T spectrum. This binning is conceptually similar to the binning created by the jet triggers at design luminosity (i.e. providing an approximately flat unweighted p_T spectrum at low p_T). Monte Carlo samples were then normalized to one another by their cross sections. For the case of ALPGEN, samples were additionally binned by the number of final state partons in the matrix-element event.

This system of re-weighting by cross section has an important implication for the variation of statistical uncertainties with p_T. Low p_T samples are given a large weight, producing a large weighted number of events in a low p_T bin. Conversely, high p_T samples are given a small weight, producing a small weighted number of events in a high p_T bin. This difference becomes important when considering the uncertainty on any variable binned in p_T. The large weighted number of events at low p_T results in an effectively small "statistical" uncertainty, and the small weighted number of events at high p_T results in an effectively large "statistical" uncertainty, even though the two p_T bins may have a similar number of *unweighted* events. Thus, the statistical uncertainties on the Monte Carlo will mimic statistical uncertainties on the data.

5.1 Event Selection

All events were included that satisfied three criteria:

- Data quality was required to be good (green) for the luminosity block holding the event [159]. This entailed stable beams from the LHC, good data quality from the pixel tracker, SCT, and TRT, and good data quality from the combined performance tracking, missing transverse energy, and jet groups. The last two requirements included a requirement of good data quality in the calorimeter. The missing transverse energy quality included a requirement of good data quality in the muon system. The solenoid was also required to be on and near nominal current.

- Events were required to pass one of the L1 jet triggers, L1_J5 or L1_J15 [164, 165, 166].

The level one jet triggers search for local maxima among the 0.1×0.1 in $\eta \times \phi$ "trigger towers" in the calorimeter with $|\eta| < 3.2$. When a local maximum is found, the total energy in a square window of size 0.4×0.4, 0.6×0.6, or 0.8×0.8 in $\eta \times \phi$, depending on the trigger configuration, is evaluated. The level one jet triggers are labeled by their transverse momentum (p_T) thresholds, which are significantly below the point at which the triggers become fully efficient. In the case of the L1_J5 trigger, the calorimeter energy is summed in a 0.4×0.4 window, and the trigger is fully ($> 99\%$) efficient for jets above $\sim$ 60 GeV. In the case of the L1_J15 trigger, the calorimeter energy is summed in a 0.8×0.8 window, and the trigger is fully ($> 99\%$) efficient for jets above $\sim$ 110 GeV. Both triggers also required coincidence with the beam pickup triggers, reducing contributions due to cosmic rays and single-beam background. The L1_J15 trigger was not prescaled during the data taking period considered in this analysis, but the L1_J5 trigger was prescaled by up to 700 (i.e. one event was kept for every 700 satisfying the trigger). During the data collection period, the jet triggers used ∼15% of the total trigger bandwidth, up to ∼60 Hz.

- Events were required to have exactly one reconstructed primary vertex with more than four tracks associated to it and with a z position within 10 cm of the geometric center of the detector. In order to reduce the contribution of secondary vertices, the vertex was required to be within five standard deviations of the luminous region in x and y. This vertex requirement further reduces the contamination from beam gas, beam halo, and cosmic ray events within the sample. It was used, in part, to avoid a cut on the timing of the cell signals on either side of the detector, since the latter requires significant energy deposit on both sides of the detector and may bias the event sample.

41 million jets in the data satisfied all of these criteria. The data quality requirement rejected ∼12% of the data, and the requirement that exactly one good vertex be reconstructed within 10 cm of the geometric center of the detector rejected ∼56% of the data. Most of these events were rejected because they contained more than one good vertex. The cut on vertex z position was 90.3% efficient when applied to the Monte Carlo simulation. Including pre-scales, about 30 nb^{-1} of data were collected with the L1_J5 trigger, and 312 nb^{-1} were collected with the L1_J15 trigger.

No event-level cut was made on reconstructed missing transverse energy. Because of the neutrinos carried along with b-jets, limiting the missing transverse energy in the event could bias the measured jet shapes. Additionally, particularly wide or narrow jets may be poorly calibrated, leading to "fake" reconstructed missing energy.

Three types of background event must be considered. If the beam scrapes against a collimator upstream of the detector, it can send a spray (mostly of muons) through the detector, called "beam halo." The beam can interact with stray gas molecules in the beam pipe throughout the cavern (called "beam-gas events"). Also, cosmic rays can deposit a sufficient amount of energy to create jets in the calorimeter.

Beam halo events are expected to occur at 50-100 Hz at design luminosity [167]. At the much lower luminosities (corresponding to much lower beam intensities) of the data included in this analysis, the rate should be suppressed by $\sim 10^4$. The requirement of a reconstructed vertex should provide further suppression by $\sim 10^2$. Beam-gas events are expected to occur at around 400 Hz at design luminosity [168], so for this analysis the rate should be below 10^{-1} Hz. The additional requirement of a vertex in the event within 10 cm of the geometric center of the detector in z further suppresses this background by about three orders of magnitude.

The rate of jets above ~20 GeV from cosmic rays in ATLAS has been measured to be a few Hz [169]. Two classes of cosmic ray events can be identified. The first class includes those cosmic rays that produced a sufficiently large energy deposition in the calorimeter to fire a level one calorimeter trigger, and the second class includes those cosmic rays in time with an event that otherwise triggered. Because all events were required to fire the level one calorimeter trigger and to be coincident with the beam pickup triggers, the first class of cosmic ray events are suppressed by at least 10^3, to be at most a few mHz. The additional requirement of a reconstructed vertex should be sufficient to make this background completely negligible. The overlap of cosmic rays with true collision events (i.e. class two) is suppressed by $\sim 10^5$ because of the average fraction of filled bunches (10^{-3}) and typical number of collisions per crossing (<1).

The total physics background from cosmic ray, beam halo, and beam-gas events is thus expected to be less than 0.01% of the sample. Additional cuts are applied to jets which further reduce the backgrounds from muons (the main contributions coming from cosmic ray and beam halo events). Thus, the backgrounds are neglected.

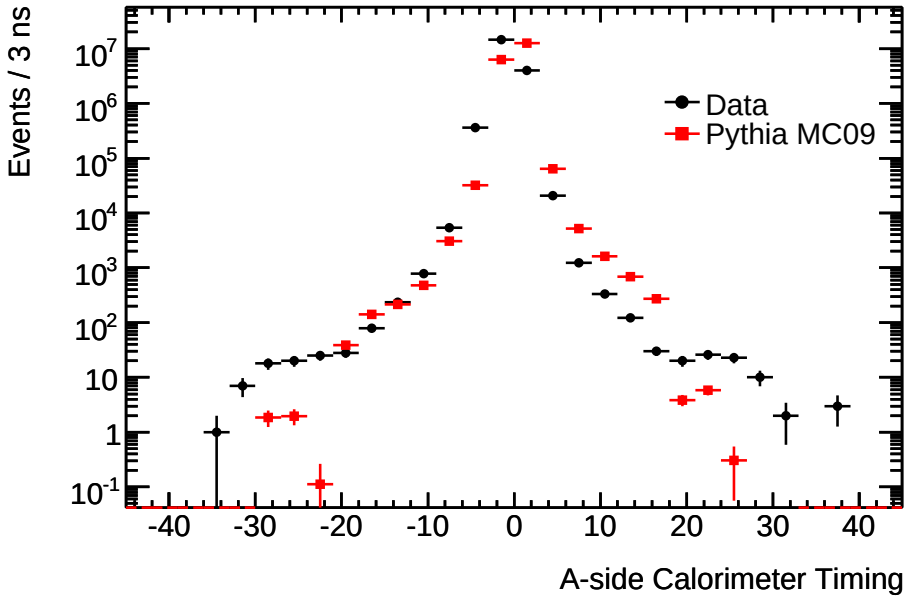


Figure 5.1: The difference in signal time between the calorimeter end-cap in positive η and negative η. Collisions should have a signal time difference near zero, whereas most backgrounds should have a signal time difference different from zero. The difference in the central region is consistent with di-jet Monte Carlo simulation done with Pythia, and the background is suppressed by more than 10^5. This estimate of the background also includes detector noise (i.e. non-physics backgrounds), which should be approximately flat in time.

An independent estimate of the background rate can be obtained from the difference in signal time in the positive η and negative η end-caps of the detector. Cell signal "time" was constructed using the pulse shape of any cell with energy at least four standard deviations above the typical noise for that cell. At least five cells were required to have a time in order to define an end-cap signal time in this manner. If the energy depositions come from proton-proton collisions, then the difference in signal time between the two end-caps should be near zero. If, on the other hand, the energy depositions come from background events (e.g. cosmic ray events, beam halo, or beam gas interactions), then the difference in signal time should be different from zero. The difference in timing on the two sides of the calorimeter is shown in Figure 5.1. The data are consistent with di-jet Monte Carlo simulation using Pythia in the central region. From the tails, the background can be estimated at $< 10^{-5}$. This estimate also includes detector noise (i.e. non-physics backgrounds), which should be approximately flat in time, and is further reduced by the jet quality cuts applied to the data.

The distributions of number of vertices per event and vertex position in z are shown in Figure 5.2. Because the Monte Carlo simulation only included a single proton-proton collision per bunch crossing, the rate of two-vertex events gives an estimate of the vertex fake rate ($\sim 10^{-3}$), and the rate of zero-vertex events gives an estimate of the vertex finding efficiency for events passing the L1_J5 or L1_J15 trigger (>99.9%). These estimates are consistent with the findings in vertex studies with the data [141]. The vertex distribution in z is marginally wider in the Monte Carlo than it is in data, because the width used was estimated from the projected beam parameters prior to the data-taking period. As jets and jet shapes were reconstructed with respect to the primary vertex, this difference was studied (see Section 5.3.7) but not corrected. The effect is below 0.5% on the final jet shape distributions.

5.2 Jet Reconstruction, Calibration, and Selection

Jet finding in this analysis comprised four steps. First, calorimeter noise was suppressed using a "topological clustering" technique. Second, jet finding was run on the remaining calorimeter signals using the Anti-k_t jet algorithm with D=0.6 (see Section 2.1 and [81]). Third, the jets were calibrated using a transverse-momentum and pseudo-rapidity dependent calibration. Fourth, the jet characteristics were checked to ensure that they were not consistent with detector noise. As this analysis only considered jet shapes, all jets in the event were treated independently (i.e. no event rejection was done when bad jets were found).

Noisy calorimeter cells can cause significant problems in jet reconstruction and may also bias jet shape measurements. There are two reasons that noise that can be problematic. One arises from cells that produce particularly large false signals during an event. These cells can be masked (i.e. removed from consideration prior to jet finding), but some cells "blink" or produce large signals only part of the time. Rather than masking these cells, since they sometimes provide useful data, "noise suppression" is used to exclude them during the reconstruction. The second problem caused by detector noise results from summing over large numbers of cells with small positive energy deposits (i.e. energy deposits from electronic noise). The simplest approach to suppress this noise is to simply omit from the reconstruction all calorimeter cells that are below a noise threshold. This type of noise

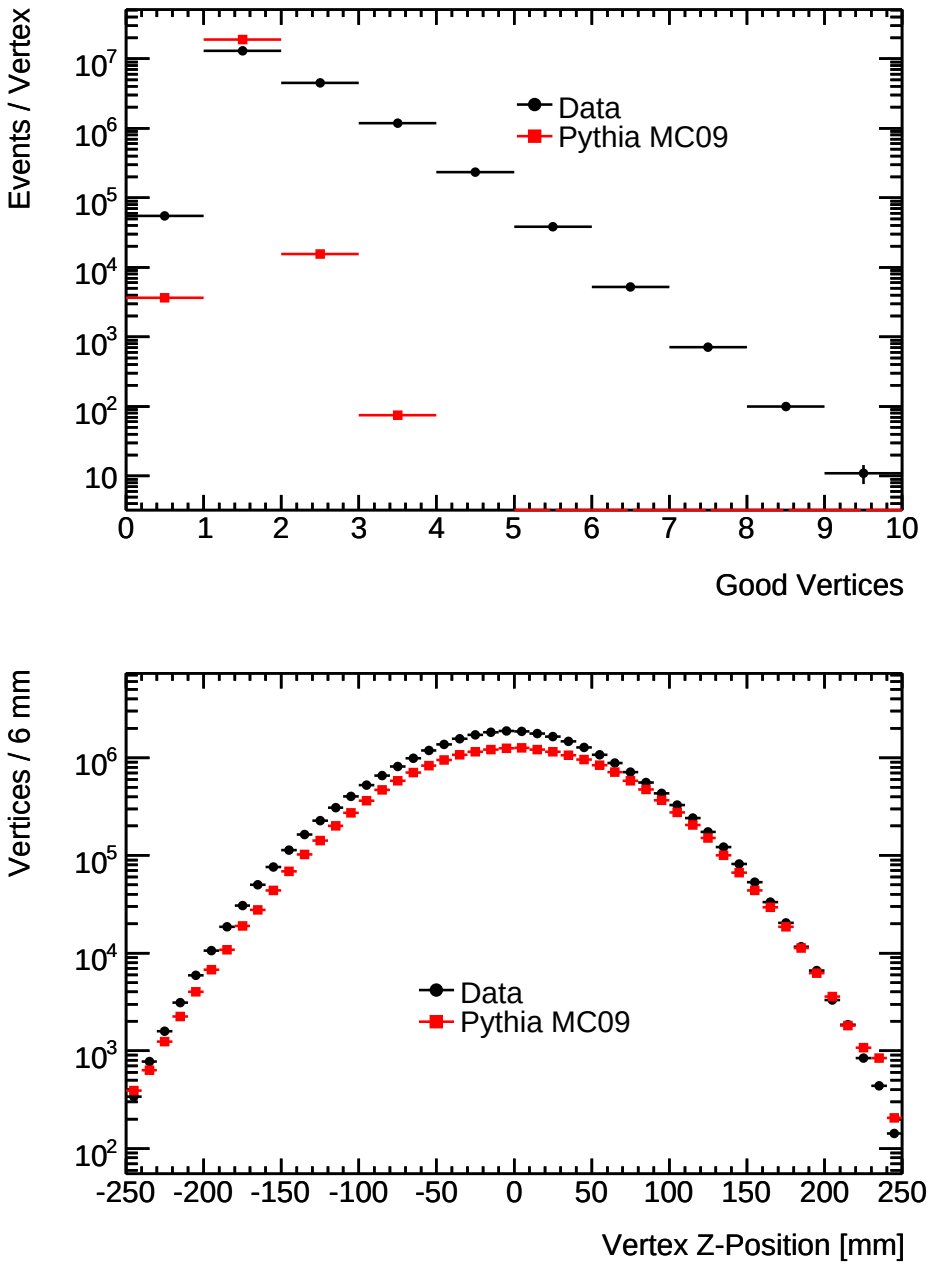


Figure 5.2: Top, the number of good vertices (vertices with more than four tracks and with a position within five standard deviations of the luminous region in x and y) in data and Monte Carlo simulation. The Monte Carlo simulation uses exactly one proton-proton interaction per event. Bottom, the z position of the vertices in data and Monte Carlo simulation. The Monte Carlo simulation used a z vertex distribution that was estimated from the expected beam parameters.

suppression, however, can cause significant problems near the edges of a shower, where the energy deposition may be small. As a result, it also may cause problems in evaluating jet shapes far from the core of the jet.

In order to reduce the effect of hot cells in isolation, ATLAS uses a calorimeter noise-suppression technique based on topological clustering of energy deposits [87]. During calibration periods without any beam, or during the abort gap, the energy of the calorimeter cells is measured. From these measurements, a mean and standard deviation are constructed. These standard deviations vary as a function of η in the detector for each layer of the calorimeter, and are shown in Figure 5.3 [128]. The noise is an order of magnitude higher in the forward region than in the barrel, but a particle of constant transverse energy would deposit an order of magnitude more energy there than it would in the barrel. A "cluster" may be seeded by any cell that is at least four standard deviations above the typical noise for that cell. Around these seeds, all cells at least two standard deviations above the noise threshold defined for that cell are included in the cluster. Finally, an additional single layer of cells is added around the cluster. It is important to note that because of the pulse reconstruction of the LAr calorimeter (see Section 4.2) it is possible for a cell to report negative energy. This final layer, therefore, may or may not increase the energy of the jet. It was added to improve the energy resolution for single pions [87].

Jet algorithms can be run directly on these (massless by construction) topological clusters, using the energy-weighted centroid of the cluster as its position. Because they include an irregular calorimeter area, using clusters may present some experimental challenges at higher luminosities. For this reason, it is desirable to work with calorimeter objects of fixed size, like calorimeter towers (see Section 4.2). These towers *always* have size 0.1×0.1 in η-ϕ. Each tower's energy may be redefined using only those cells included in a topological cluster, and jet algorithms may be run on these "topo-towers." In this case, the topological clustering is only being used as a noise-suppression technique. The jet finding is run on the towers, which are made massless using the Snowmass summation technique [72]. In both cases, jet finding is done in $y-\phi$ coordinates (i.e. jet masses are allowed and taken into account) and jet four-momenta are defined as the four-vector sum of the momenta of the constituents (whether clusters or towers).

The ATLAS calorimeter is non-compensating, meaning that the energy response for charged pions differs significantly from the response for photons and electrons of the same

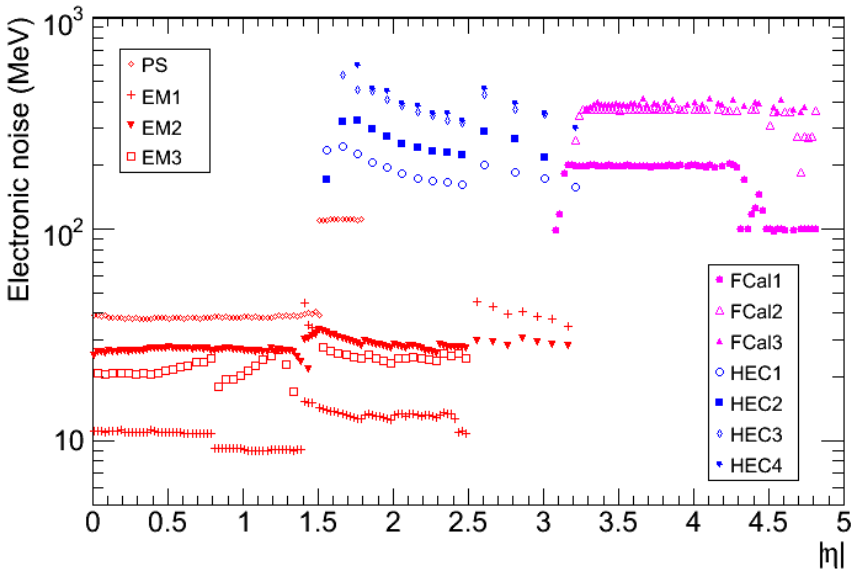


Figure 5.3: The noise thresholds in the calorimeter as a function of detector η for each layer of the calorimeter. The presampler (PS) is in front of the electromagnetic (EM) calorimetry up to $\eta = 1.8$ The hadronic end-cap calorimeter (HEC) and forward calorimeter (FCAL) have the most noise on average. The first layer of the FCAL acts as an EM calorimeter.

energy. In order to take into account this non-compensation, it is necessary to correct the hadronic component of a jet's energy. Of course, it can be difficult to evaluate exactly how much of a jet was "hadronic." There are several methods for doing so (called "calibration techniques") that have been evaluated by ATLAS [87]. The simplest technique simply scales the measured jet energy to account for the average hadronic fraction of a jet of similar energy and rapidity. Any more complicated methods have been reserved until sufficient data have been collected that those methods might be validated *in situ*. For improving the jet energy or transverse momentum resolution, a more complex calibration will be desirable.

Jet calibration takes two steps. First, before any noise suppression is applied, each cell is corrected to the electromagnetic scale (EM scale). This correction takes into account the sampling fraction of the calorimeter and any shaping effects in the electronics. The second step of calibration takes into account the difference in response between hadrons and electrons. For the jets considered in this analysis the calibration technique applied was "numerical inversion" [170, 171]. Corrections were derived entirely from Monte Carlo simulation. After correction to the EM scale, the jet energy response is defined as the ratio

of energies of the calorimeter jet at the EM scale and the hadron-level jet. The jet energy response is fit as a function of jet p_T for bins with a width of 0.1 in $|\eta|$[1]. The calibrated energy for each jet is then defined as $E_{\text{calib}} = E_{\text{EM}}/R\,(p_T, \eta)$. After calibration, the energy of a typical jet is increased by ~40%. The distributions of transverse momenta of jets passing all cuts before and after hadronic calibration are shown in Figure 5.4.

Because of the fitting procedure, the averaging over each bin, and the possible differences in kinematics between the sample used to derive the correction and the sample to which the correction is applied, this calibration is imperfect. The systematic uncertainty on this calibration was below 9% (below 7% for jets with $y < 0.8$), and varied as a function of jet pseudo-rapidity and transverse momentum [170]. This uncertainty was derived using Monte Carlo simulation and was cross-checked using test-beam results and single particle analysis in the collision data. The uncertainty for central jets ($y < 0.3$) as a function of jet transverse momentum is shown in Figure 5.5. In order to take into account contributions from other proton-proton collisions in the event, the systematic uncertainty was derived as a function of the number of good vertices in the event.

This calibration scheme provides a single correction for each jet, rather than correcting the individual jet constituents. As such, jet shapes calculated using EM-scale objects are identical to those calculated using calibrated objects. For simplicity, jet shapes are all calculated here using EM-scale constituents (towers, unless otherwise specified). Some of the jet calibration techniques ATLAS has studied involve applying different weights to different clusters or cells inside the jet. If one of these calibrations is applied, then to be completely self-consistent the calibrated constituent could be used to construct the jet shapes. However, as has been noted earlier, there is no need to construct jet shapes from exactly those objects that were used by the jet finding algorithm. For this same reason, it still interesting to consider either "calibrated constituents" or constituents at the EM scale, even when using these calibration schemes. These issues will be explored further in Section 5.6.2.

After jet finding, a correction is applied to account for the origin of the jet being displaced from the geometric origin of the detector. A primary vertex is selected in the event using the criteria described in Section 5.1. The beam spot is used if no primary vertex can be found,

[1] η is the relevant quantity here because response varies most strongly as a function of the detector features, including cracks and transitions between technologies.

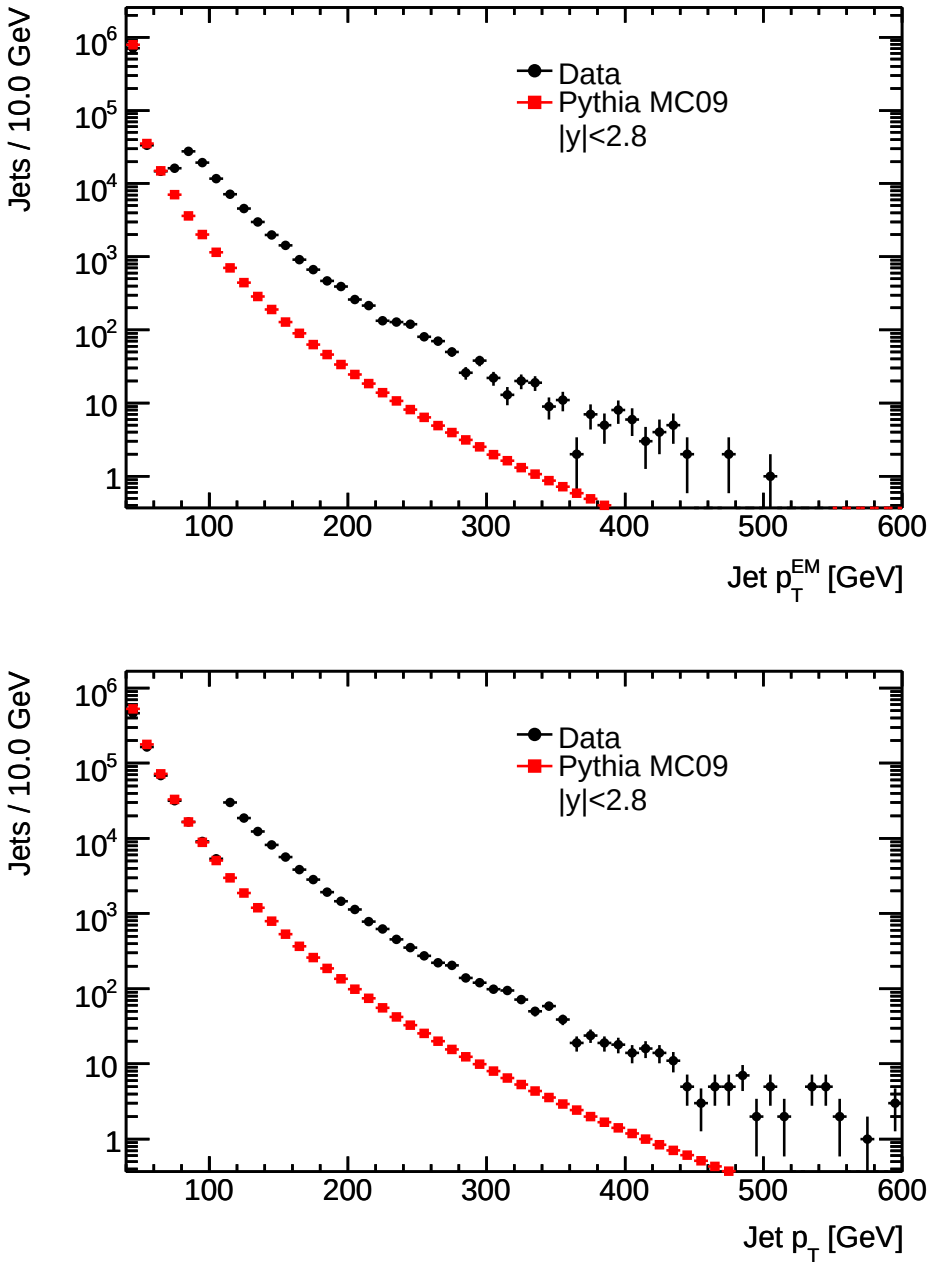


Figure 5.4: Transverse momentum spectrum of jets passing all quality cuts with $|y| < 2.8$ and 40< p_T <600 GeV calibrated transverse momentum before (top) and after (bottom) hadronic calibration. The left-most bin of the uncalibrated transverse momentum spectra include jets with calibrated p_T above 40 GeV but uncalibrated p_T below 40 GeV. The data have a somewhat harder p_T spectrum than the Monte Carlo simulation. The discontinuity at 110 GeV is a result of the different triggers being used below (L1_J5) and above (L1_J15) that threshold. The Monte Carlo simulation distribution is normalized to have the same number of good jets as the data.

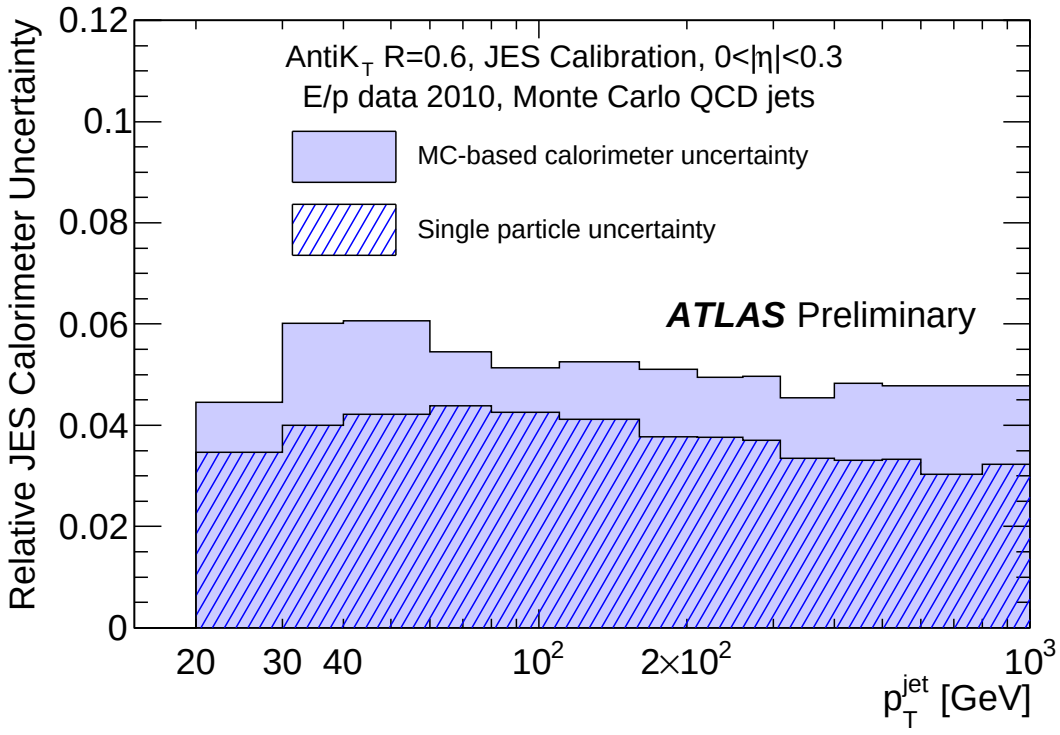


Figure 5.5: The jet energy scale systematic uncertainty for central jets ($y < 0.3$) as a function of jet transverse momentum. The uncertainty is shown from both a pure-Monte Carlo analysis (solid fill) and from an *in situ* single particle analysis. The larger Monte Carlo driven uncertainty is used in this analysis.

and the vertex with the highest sum of the p_T^2 of all tracks associated to it is used in the case of multiple vertices. The constituents' positions are then recalculated using this modified origin, and the constituents are re-summed to calculate the new jet momentum. Although the projected "size" of the jet may change because of this correction, no modification to the constituents of the jet is allowed at this stage (i.e. constituents are not allowed to fall into or out of the jet "cone," nor is jet finding re-run). This vertex displacement is also taken into account in calculating the jet shapes.

Four jet cleaning cuts were applied to reduce contributions from calorimeter noise, mis-measurements, and physics backgrounds (e.g. cosmic rays or beam halo). Although masking was applied to remove particularly noisy cells, some cells that were occasionally noisy, "blinking," or noisy during collision events (as appeared to be the case with some cells in the hadronic end-cap calorimeter [159]) were not masked. Therefore, jets from calorimeter noise need to be removed. These cleaning cuts made use of five variables:

- jet "time," the E^2-weighted sum of times from all cells in the jet;
- jet "EM fraction" (EMf), the fraction of jet energy at the EM scale deposited in the electromagnetic calorimeter;
- jet "HEC fraction" ($HECf$), the fraction of jet energy at the EM scale deposited in the hadronic end-cap calorimeter;
- jet "N90," the minimum number of cells in the jet containing 90% of the jet's energy;
- jet "Quality" (Q), the fraction of the jet's energy coming from cells with a signal quality above 4000. The signal quality can be thought of as a χ^2 between the measured signal and a predicted signal with the same time and amplitude.

Both time and quality are computed only using those cells with recorded energy at least two standard deviations above their mean noise.

Using these cleaning variables, the following cuts were applied, following References [159, 172, 173]:

- Jets were required to have *either* an EM fraction less than or equal to 95% *or* quality less than 0.8. This cut removed noise in the electromagnetic calorimetry.
- Jets were required to have *either* a HEC fraction greater than or equal to 20% *or* N90 greater than four. This cut removed noise in the hadronic end-cap calorimeter.
- Jets were required to have jet quality less than one minus the HEC fraction ($Q < 1 - HECf$). This cut further removed noise in the hadronic end-cap calorimeter.
- Jets were required to have a time within 50 ns of the collision. This cut removed cosmic rays and detector noise.

Because the vast majority of the events at the LHC are multi-jet events, there is a natural cross section based suppression of leptonic events. Therefore, no identification or removal of leptons is done. This is also consistent with the previous studies from the Tevatron [96, 97, 98]. The rate of isolated leptons is suppressed relative to jet production by at least three orders of magnitude [128].

The distribution of the jet HEC fraction as a function of N90 is shown in Figure 5.6. The cut, removing jets with HEC fraction less than 20% and N90 less than five, is 99% efficient

in data, and in Monte Carlo simulation is well over 99.99% efficient. In early studies at LEP, N90 was evaluated as a discriminant between quark- and gluon-jets (albeit in a calorimeter with larger granularity cells) [101]. Therefore, one must be particularly careful to not apply a much stronger cut with this variable. The high efficiency provides some confidence that only noise is being rejected.

The distributions of the jet signal quality as a function of the jet EM fraction in data and Monte Carlo simulation are shown in Figure 5.7. The jet quality is very poorly modeled in Monte Carlo simulation. Thus, although the cut is 100% efficient in Monte Carlo simulation, the efficiency cannot be estimated there. However, the cut is over 99.99% efficient in data.

The distribution of the jet HEC fraction and jet quality are show in Figure 5.8. The core of the distribution of HEC fraction is fairly well modeled in the Monte Carlo. However, because of the poor modeling of the jet quality, the efficiency of the cut cannot be estimated from Monte Carlo simulation. The cut is over 99.9% efficient in data as well.

The distributions of jet time in data and di-jet Monte Carlo simulation are shown in Figure 5.9. This cut is also 99.99% efficient in both data and Monte Carlo simulation. The distribution is clearly wider in the data than in the Monte Carlo simulation, which is a result of the detectors and triggers not being completely timed in (see Section 4.5). In the future, when the detector and triggers are better timed, this cut might be made considerably tighter. Once most bunches of the LHC are filled, the cosmic ray and beam halo backgrounds may be several orders of magnitude larger. At that time, a tighter timing cut will be more useful to reject cosmic ray and beam halo background.

Distributions of jets in η and ϕ after all quality cuts have been applied are shown for both data and Monte Carlo simulation in Figure 5.10. The "cold" regions of the detector correspond to dead optical transmitters in the LAr calorimeter and dead modules in the tile calorimeter (see Section 4.2). The differences between the distributions correspond to detector issues that arose during the data-taking period and, therefore, were not present in the simulation. Each dead or masked cell in the calorimetry, excluding ones inside or neighboring a region with a dead optical transmitter in the LAr calorimeter, is given the average energy of its neighbors. The fraction of jet energy deriving from this correction is also shown in Figure 5.9. The agreement between the data and the Monte Carlo simulation is good.

Figure 5.11 shows the number of good jets per event after all jet quality cuts have been

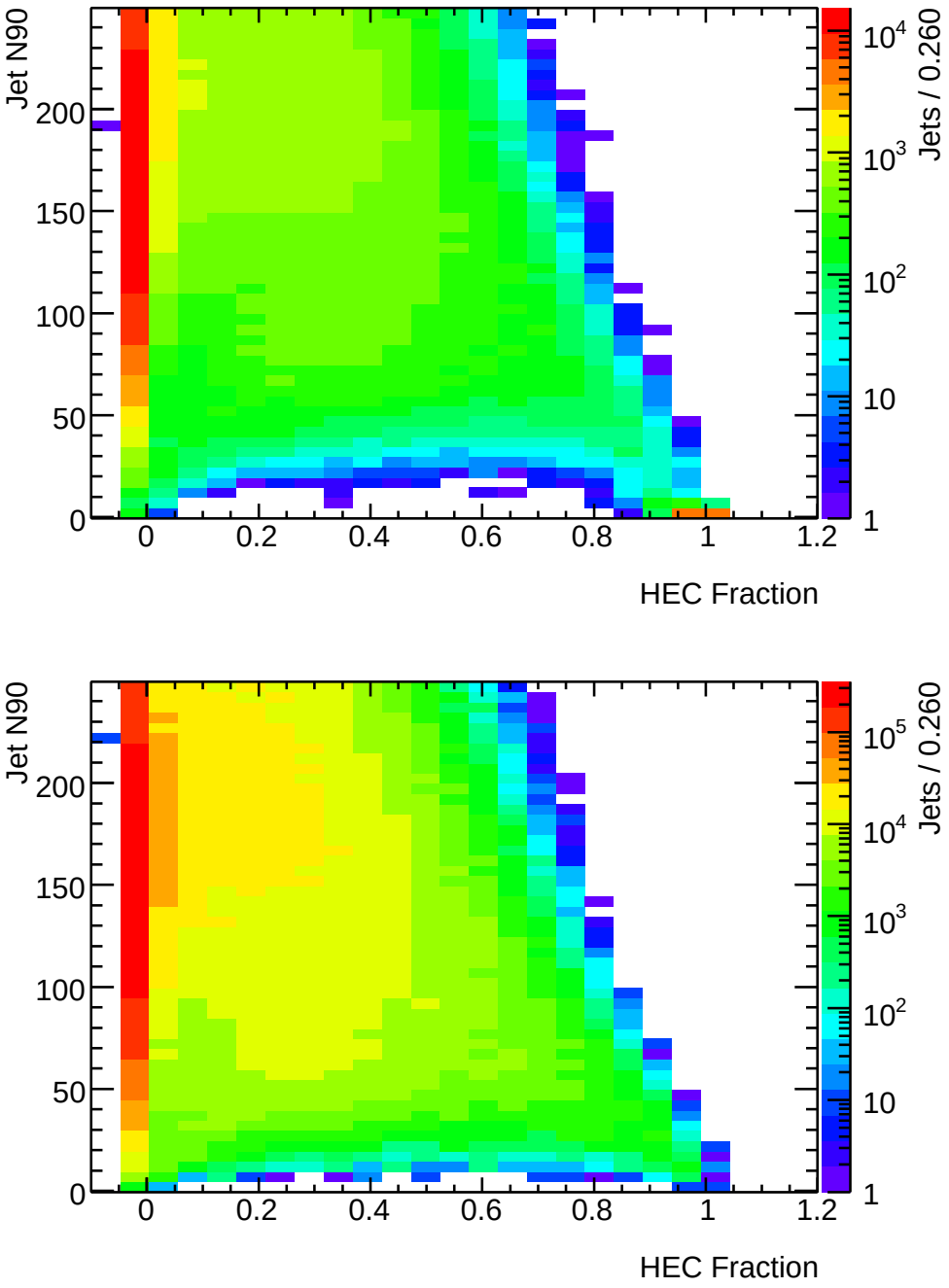


Figure 5.6: The distribution of jet HEC fraction as a function of N90 in data (top) and Monte Carlo simulation (bottom) for jets passing all other quality criteria. The large number of jets with HEC fraction near zero are mainly central jets ($y < 1.4$).

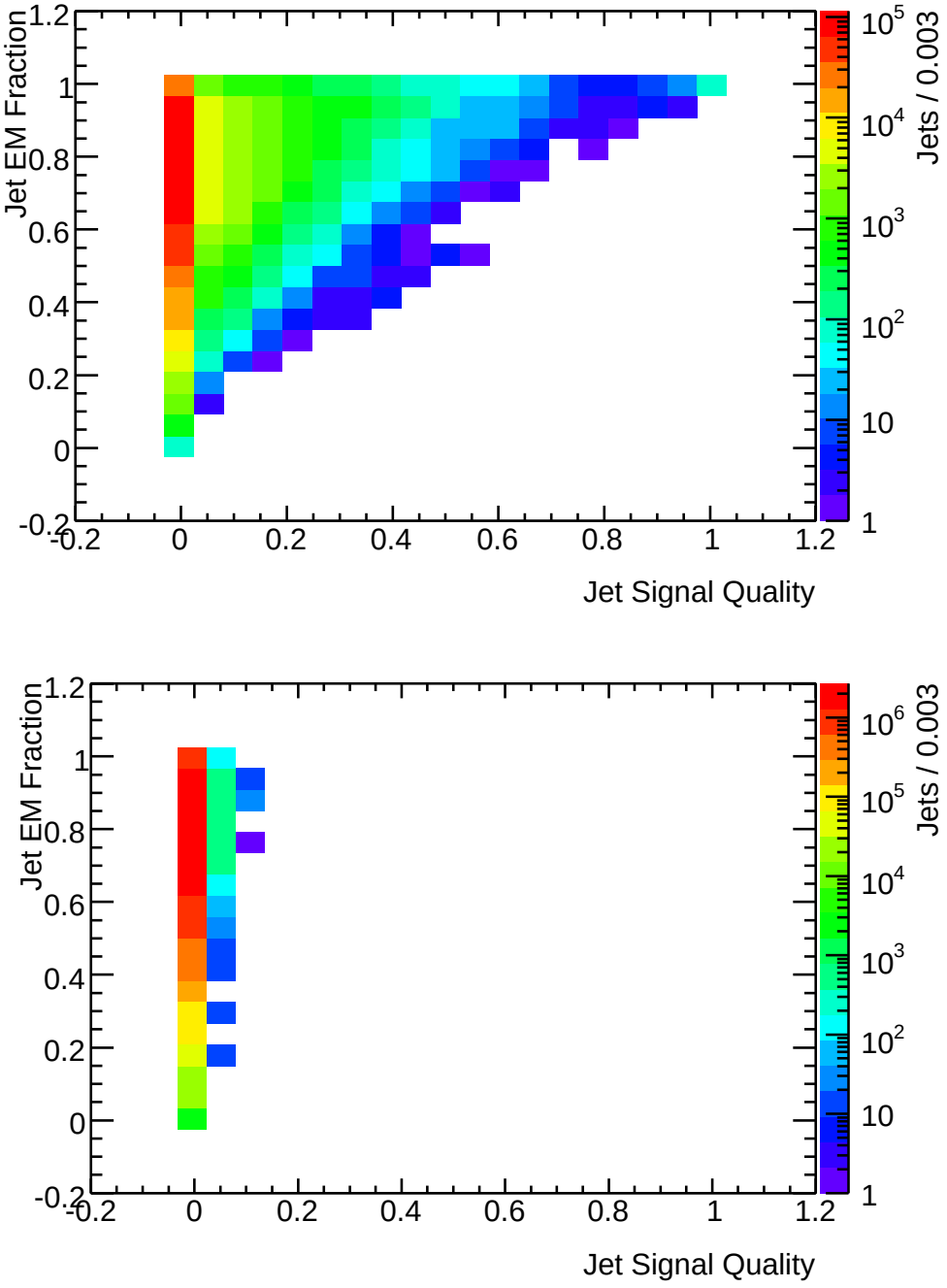


Figure 5.7: The distribution of jet signal quality as a function of jet EM fraction in data (top) and Monte Carlo (bottom) for jets passing all other quality criteria.

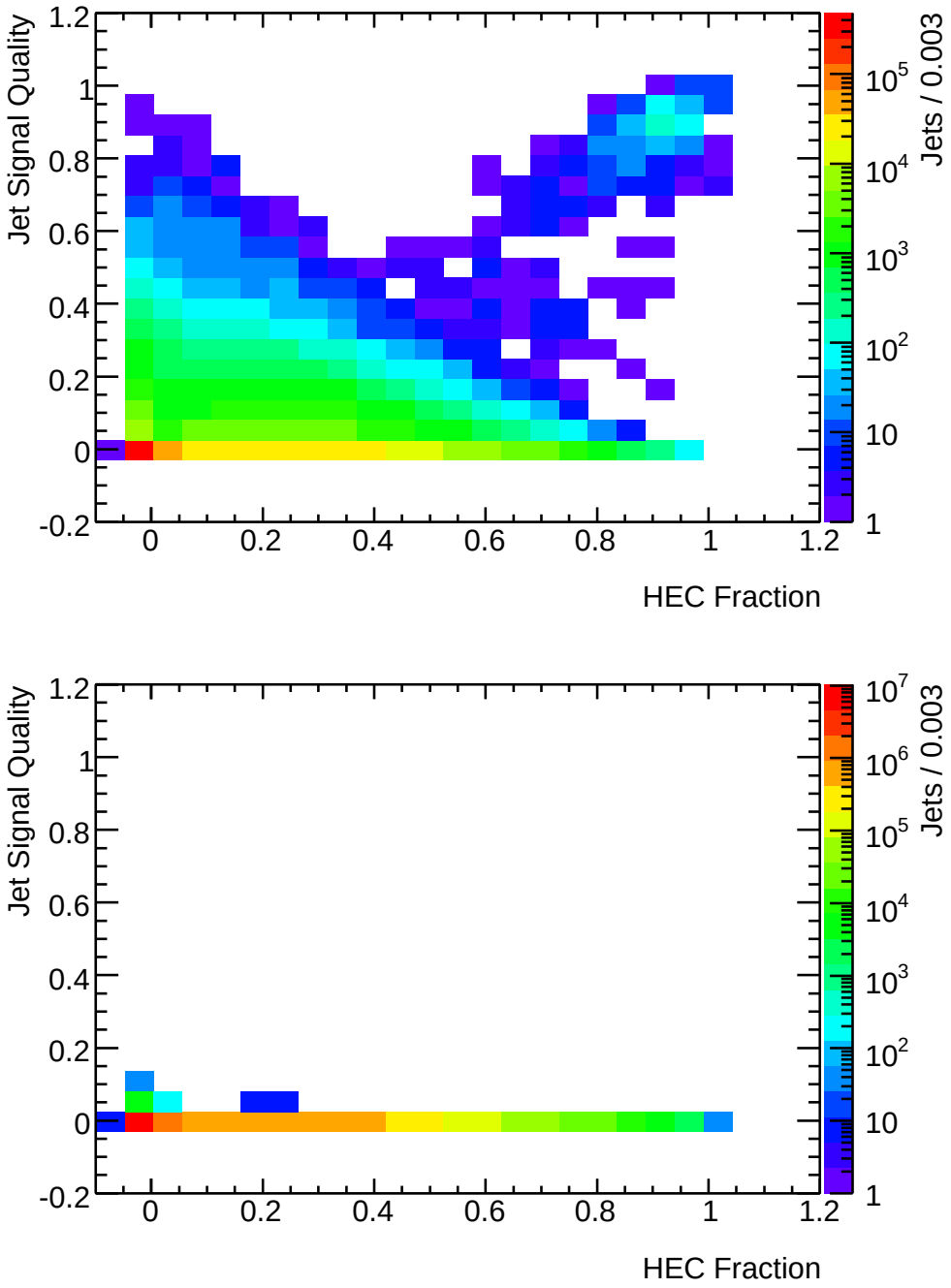


Figure 5.8: The distribution of jet signal quality as a function of the jet HEC fraction in data (top) and Monte Carlo (bottom) for jets passing all other quality criteria.

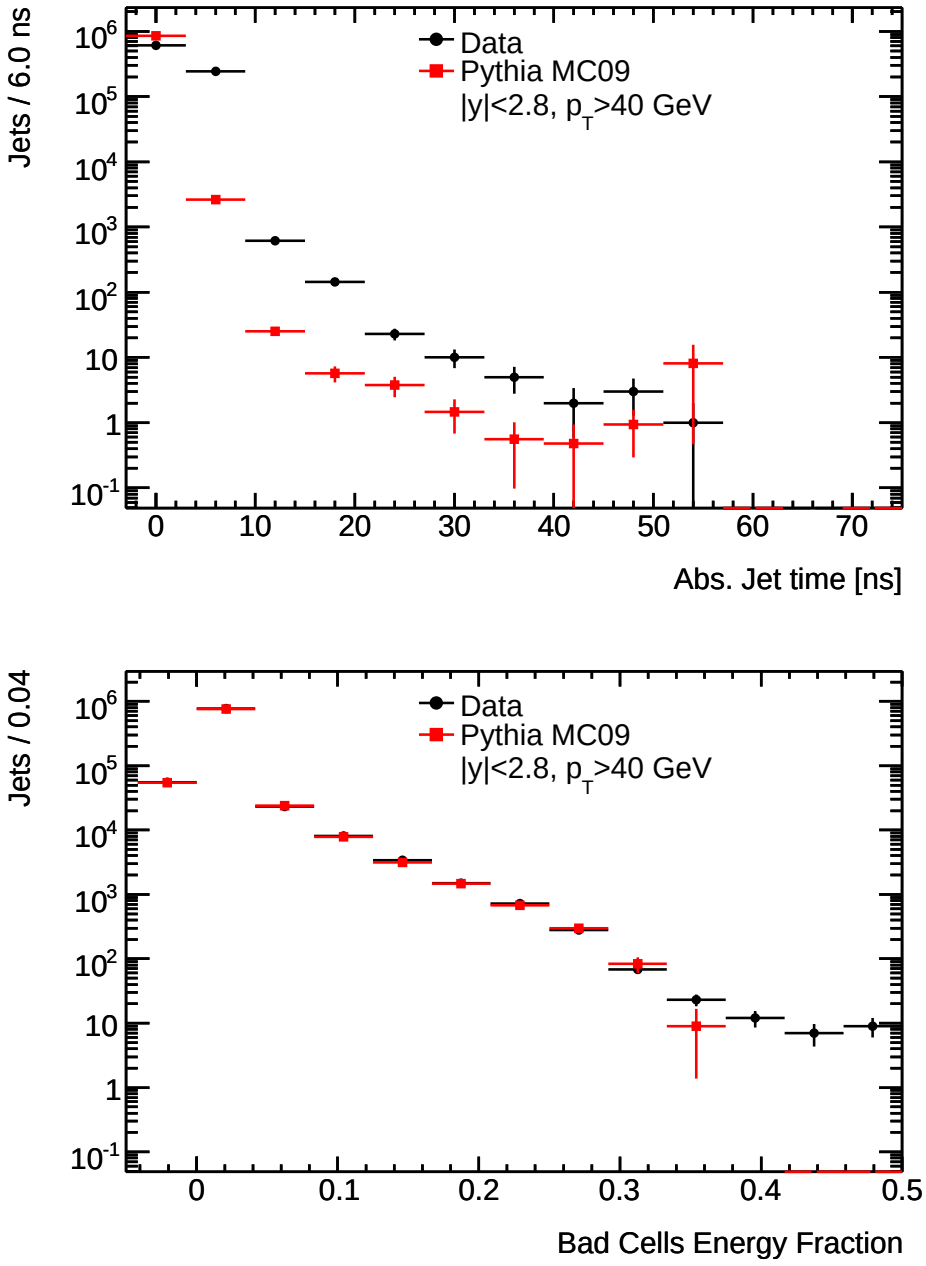


Figure 5.9: Top, the distribution of the absolute value of jet time for jets passing all other quality criteria. The agreement between the data and the Monte Carlo simulation is poor, because the detector has not been sufficiently timed in (see Section 4.5). As the detectors are timed in, the agreement is expected to improve, at which point the cut (currently 50 ns) may be tightened. Bottom, the fraction of jet energy deriving from bad calorimeter cells. The agreement between data and Monte Carlo simulation is good.

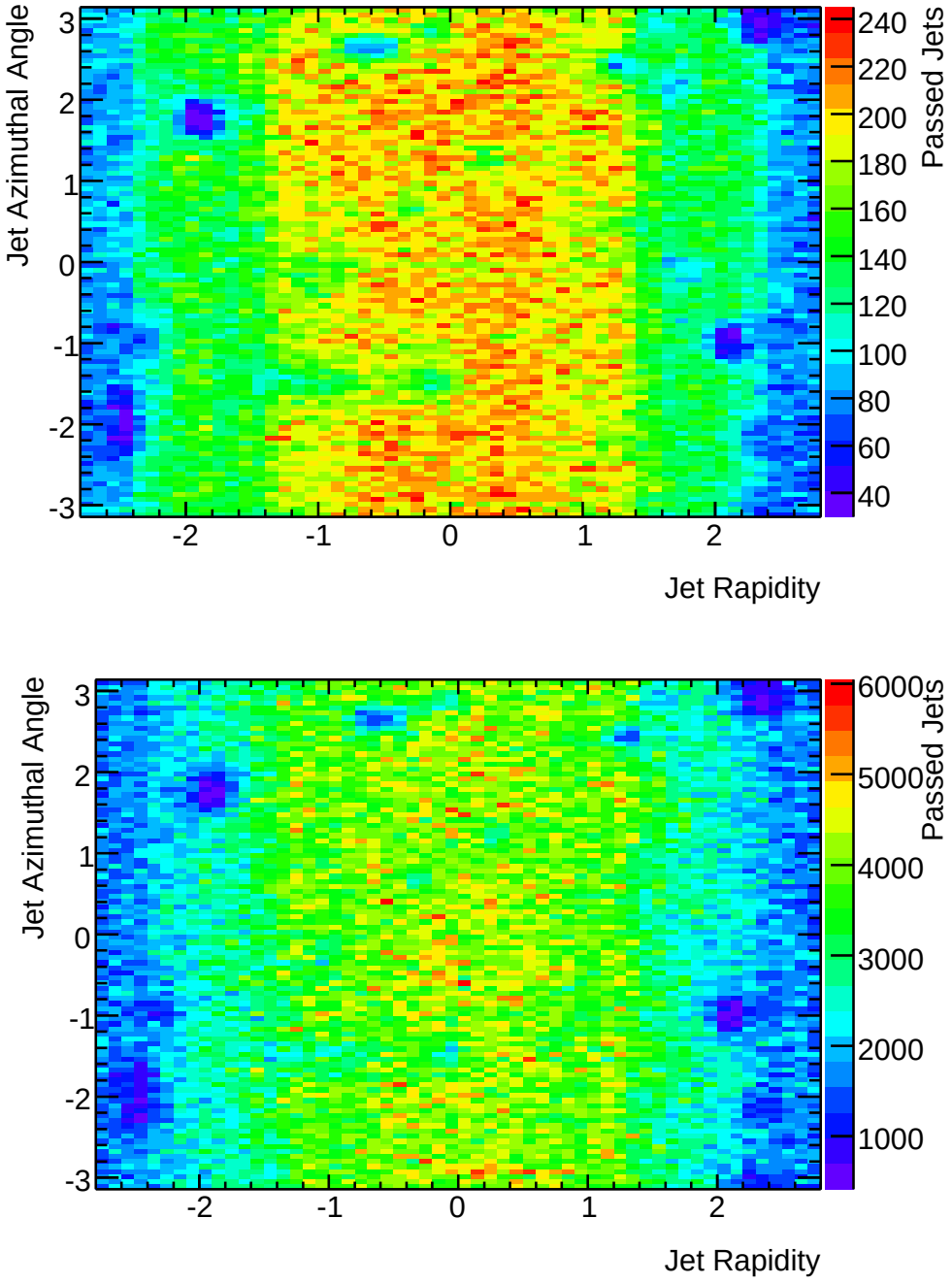


Figure 5.10: Distribution in η and ϕ of jets passing all quality cuts with $|y| < 2.8$ and with $p_T > 40$ GeV in data (top) and Monte Carlo simulation (bottom).

applied. The distributions are shown normalized to the same number of good jets. The Monte Carlo simulation follows the data well in the bins with one or more jets. Most of the single-jet events are di-jet or multi-jet events in which only one jet passed the kinematic selection. The Monte Carlo simulation has a significant number of events in the zero-jet bin, which are predominantly jets with $p_T < 40$ GeV but which fired the L1_J5 trigger. In this bin there is significant disagreement between the data and the Monte Carlo simulation, which may arise from a poor modeling of low p_T jets (in particular the contributions from the underlying event), incomplete modeling of the detector noise in the simulation, and inaccurate modeling of the level one calorimeter trigger's analog readout chain.

If events have been cleaned well, then the majority of events with exactly two good jets (di-jet events) should be back-to-back. Figure 5.12 shows the distribution of $\Delta\phi$, the difference in azimuthal angle, between a jet and the nearest jet, the closest in $\Delta R \equiv \sqrt{\Delta\phi^2 + \Delta\eta^2}$ (normalized to the number of events in the di-jet data). A sharp peak can be seen at π for all events, and the peak is further sharpened in di-jet events. Figure 5.12 also shows the opening angle (ΔR) between each jet and the nearest jet in $\eta - \phi$ space. Again, a peak can be seen at π. A smaller secondary peak is visible at $\Delta R \sim 0.6$, approximately the radius of the jet, where two jets have been reconstructed immediately next to one-another. The secondary peak is typically caused by radiation from the parton that forms a jet. If the radiation is sufficiently high-energy and sufficiently far from the core of the original jet, it will be distinguished by the jet algorithm as an independent jet. As expected, the secondary peak disappears in the di-jet events.

It may be advantageous to consider only isolated jets, since two jets reconstructed nearby one another may indicate a single, very broad jet that has been split by the jet algorithm. However, any isolation criterion is also a restriction on the shape of the jet, since it removes the possibility of nearby soft radiation. Therefore, no jet isolation cut is applied for the primary study, and isolation is only considered as a secondary effect. The effect of such an isolation criterion on jet shape distributions is explored further in Section 5.6.3.

For good jets passing all quality criteria, the distribution of minimum and maximum distance to a trigger jet is shown in Figure 5.13. The distribution of minimum distance shows a sharp peak near 0, meaning that the vast majority of good jets also produced a trigger jet. There is a secondary peak at π, typically from di-jet events where only one of the two back-to-back offline jets produced a trigger jet. The distribution of maximum

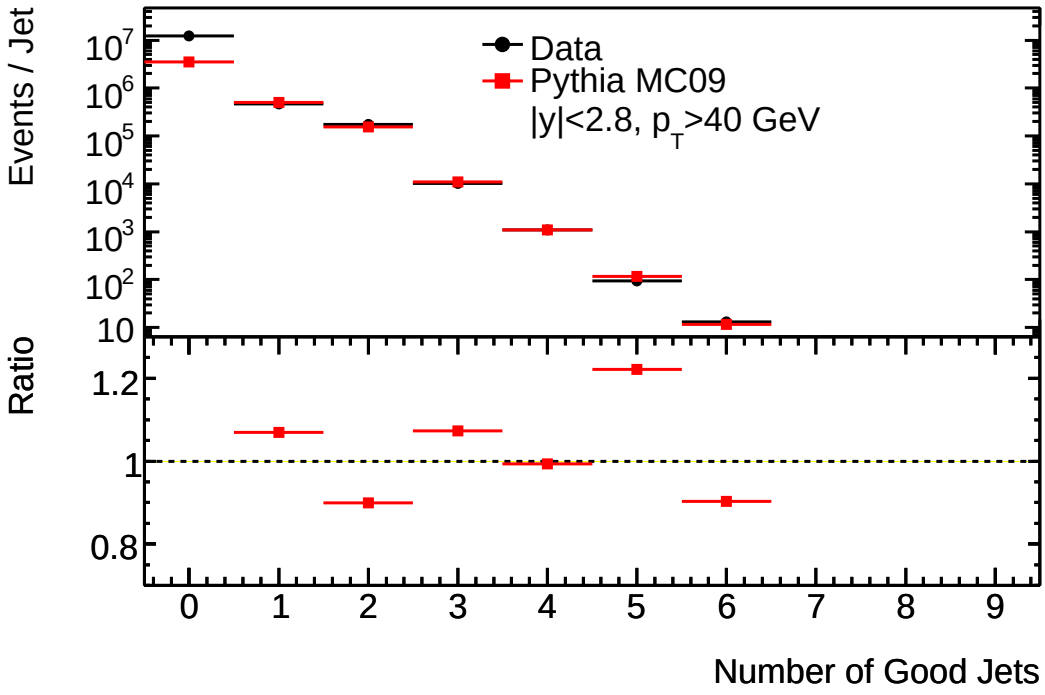


Figure 5.11: Number of jets per event passing all quality cuts with $|y| < 2.8$ and $p_T > 40$ GeV. The samples are scaled to the same number of good jets, and the ratio is defined as the jets in Monte Carlo simulation divided by the number in data.

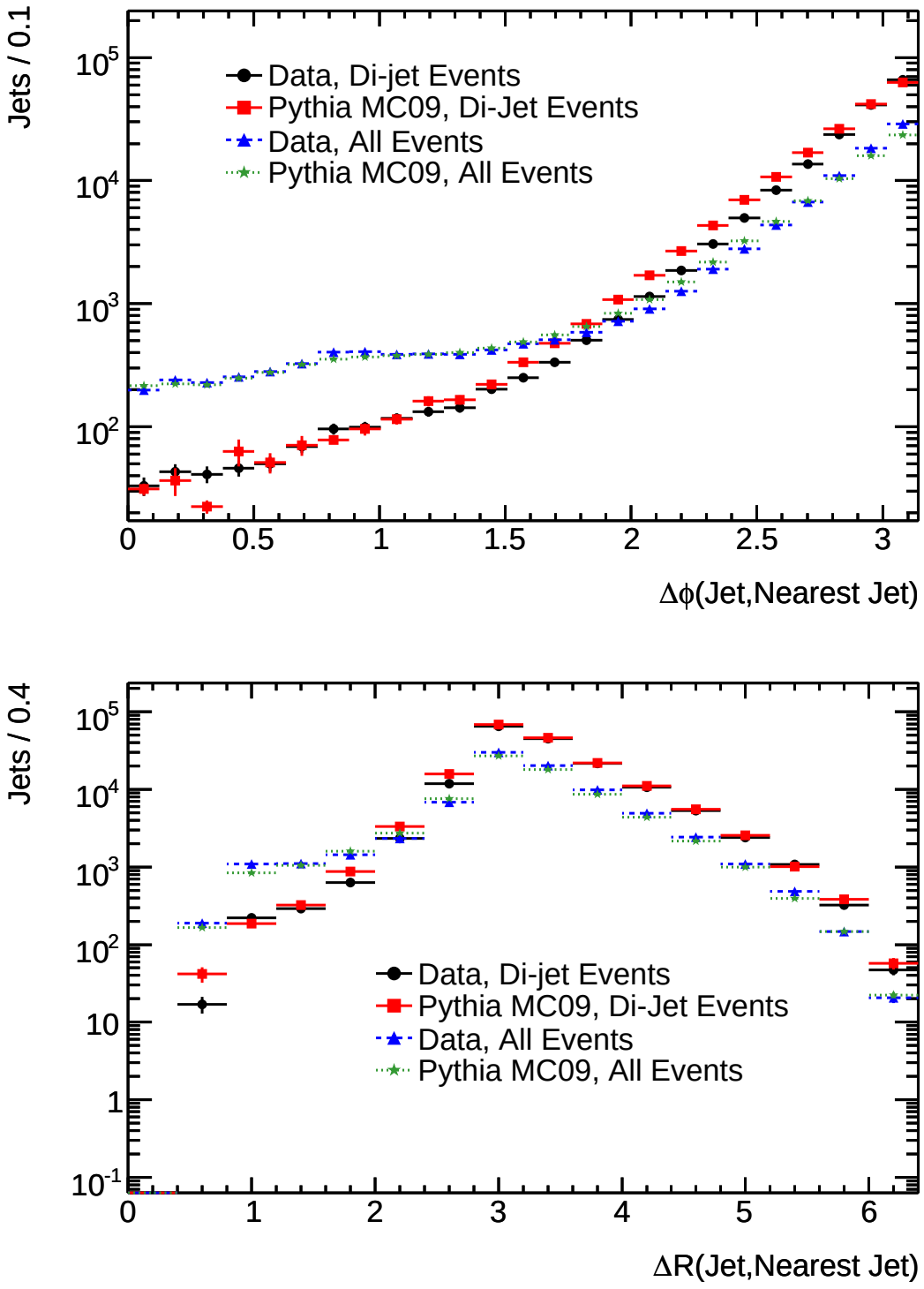


Figure 5.12: Top, azimuthal angle difference, $\Delta\phi$, between a jet and the nearest jet in $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$ in all events and di-jet events. Bottom, ΔR between a jet and the nearest jet. All samples are normalized to the number of good jets in the di-jet data.

distance shows a sharp peak at π, due to di-jet events where the away jet (also) produced a trigger jet. The L1_J5 jet trigger is ∼95% efficient at 40 GeV and is ∼99% efficient at 60 GeV. As the p_T distribution is dominated by jets between 40 and 60 GeV, a significant fraction of jets lie below the full turn-on point of the trigger. Also, the jet triggers are based on a different reconstruction algorithm and different data than the offline reconstruction. Even the largest jet trigger window, for example, corresponds to a jet with a square area of 0.64, only about 60% of the area of an Anti-k_t jet with D=0.6. The use of triggering jets could create a bias towards narrow jets. Conversely, requiring jets to be away from trigger jets could create an anti-bias. These issues are explored further in Section 5.4.2.

In order to restrict the study to the best-understood region of the calorimeter, jets were required to have $|y| < 2.8$. The jet shapes studies were restricted to jets above 40 GeV, the well-calibrated region where the jet trigger efficiency is high. For jets below 110 GeV, the L1_J5 trigger was required to have fired. For jets above 110 GeV, the L1_J15 trigger was required to have fired. This division ensured that the trigger efficiency for jets above 60 GeV was above 99%, while including as much integrated luminosity as possible. The jet reconstruction efficiency is expected to be near 100% above 30 GeV [174]. Four bins were used in rapidity. The most central bin includes a completely homogenous region of the calorimeter. The second and third bins include the transition regions between the tile calorimeter barrel and extended barrel and between the LAr barrel and end-cap, respectively. The fourth bin is entirely within the LAr end-cap. Tables 5.1 and 5.2 show the number of jets remaining after all cuts in each p_T and y bin in data and Monte Carlo simulation. The normalization between data and Monte Carlo simulation is arbitrary.

5.3 Jet Shape Unfolding

In order to account for differences between the jet shapes measured in data (using experimental quantities) and those extracted directly from stable hadrons in Monte Carlo, it is necessary to apply some "unfolding" to the data. This unfolding takes the form of a bin-by-bin correction, applied to the data, to extract the jet shapes before any smearing from experimental effects. Some unfolding procedure is necessary in order to compare these results to those of previous experiments. Complete unfolding analysis will be shown for two jet shapes: $\Psi(0.3)/\Psi(0.7)$ and $\rho^a(r)/\Psi(0.7)$. The choice of $\Psi(0.3)/\Psi(0.7)$ as a p_T-

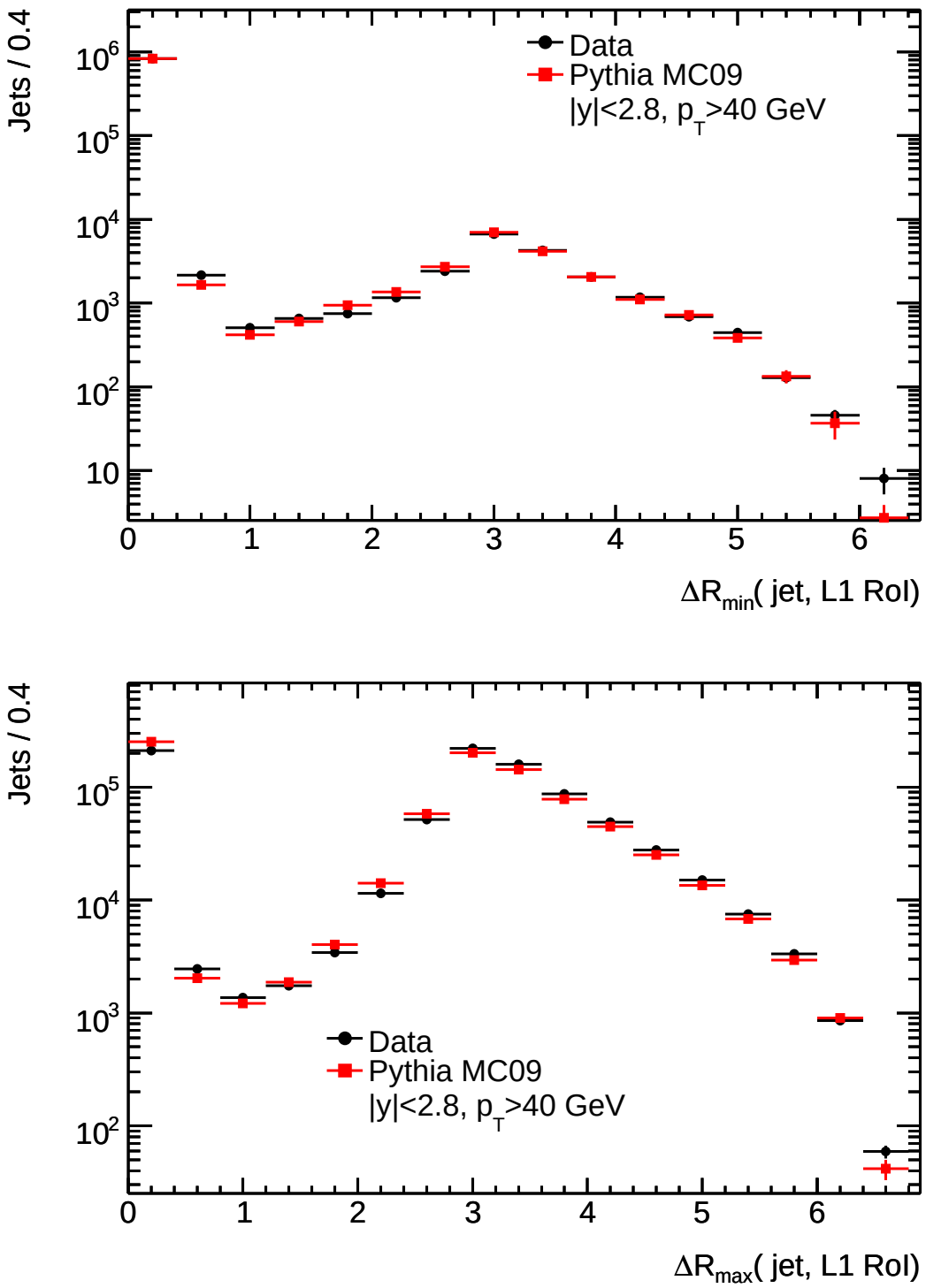


Figure 5.13: Minimum (top) and maximum (bottom) distance in ΔR to a trigger jet for jets passing all quality cuts.

Jet p_T	$\|y\|$ <0.8	0.8< $\|y\|$ <1.2	1.2< $\|y\|$ <2.1	2.1< $\|y\|$ <2.8	Total
40-60 GeV	226461	108596	191632	103227	629916
60-80 GeV	37095	17675	30742	14992	100504
80-110 GeV	11677	5439	9707	4147	30970
110-160 GeV	29830	13211	22914	8898	74853
160-210 GeV	4644	2000	3509	1022	11175
210-260 GeV	1122	485	707	173	2487
260-310 GeV	378	152	223	32	785
310-400 GeV	198	71	110	16	395
400-500 GeV	44	16	21	1	82
500-600 GeV	15	3	5	0	23
Total	311464	147648	259570	132508	851190

Table 5.1: Jets remaining in each p_T and y bin after all cuts in 7 TeV proton-proton collision data.

dependent jet shape is purely historical. It is used here to demonstrate that the "core" of a jet carried a larger fraction of the transverse momentum at higher p_T than the "tail." However, other selections may be useful for different purposes. Two alternate variables will be explored in Section 6.1.

Here the jet shapes are normalized to $\Psi(0.7)$ rather than $\Psi(0.6)$. The jet "cone" of the Anti-k_t jet algorithm with D=0.6 is close to 0.6, but two factors led to the choice of 0.7 for normalization. First, historical jet shapes have been measured using 0.7 cone jets, and using the same denominator provides some additional consistency. Second, the Anti-k_t jet algorithm can be visualized as putting a minimum of energy near the edge of the jet (see Section 2.1). Because of this, small changes in the jet axis can produce significant distortions to the jet shape immediately at the edge of the jet, as will be shown in subsequent sections. These distortions typically take the form of a significant increase in $\Psi(0.6)$ and a corresponding decrease of $\Psi(0.7)$. Thus, by normalizing shapes to $\Psi(0.7)$, these fluctuations are suppressed.

To evaluate the necessary unfolding correction, jet shapes were evaluated in Monte Carlo before and after the full detector simulation. The simulation applied all the anticipated detector effects, including:

- charged particles bending in the magnetic field in the inner detector
- particle interactions with material in the inner detector and cryostat (all the material

Jet p_T	$\|y\|$ <0.8	0.8< $\|y\|$ <1.2	1.2< $\|y\|$ <2.1	2.1< $\|y\|$ <2.8	Total
40-60 GeV	5.26×10^6	2.48×10^6	4.32×10^6	2.15×10^6	1.42×10^7
60-80 GeV	825000	379000	636000	282000	2.12×10^6
80-110 GeV	249000	113000	181000	70000	614000
110-160 GeV	63900	28200	42500	14100	149000
160-210 GeV	9710	4160	5780	1520	21200
210-260 GeV	2260	924	1230	236	4650
260-310 GeV	676	270	336	46.0	1330
310-400 GeV	325	126	143	11.8	606
400-500 GeV	72.7	26.9	25.5	0.834	126
500-600 GeV	14.0	4.80	3.72	0.0340	22.5
Total	6.41×10^6	3.01×10^6	5.19×10^6	2.52×10^6	1.71×10^7

Table 5.2: Weighted number of jets remaining in each p_T and y bin after all cuts in Monte Carlo simulation.

before the sampling portion of the calorimeter begins)

- hadronic and electromagnetic showering within the calorimeter
- calorimeter noise

To each reconstructed observable, a correction factor is applied that is derived purely from Monte Carlo. The correction factor C_i, for bin i, is simply:

$$C_i = \frac{O_i^{MC}}{R_i^{MC}} \tag{5.1}$$

where R is the reconstructed observable, O is the hadron-level observable, and MC denotes Monte Carlo simulation. Using this correction, the predicted hadron level distribution is:

$$O_i^{\text{predicted}} = C_i \times R_i \tag{5.2}$$

where R_i is the measured observable in the data. Using this procedure, unfolding the Monte Carlo used to derive the correction factors must exactly reproduce the hadron-level distributions.

Unfolding corrections were derived using Pythia with the ATLAS MC09 tune [63]. The unfolding corrections are shown in Tables 5.3 and 5.4 and are generally near one. The largest unfolding correction for $\Psi(0.3)/\Psi(0.7)$ is less than 5%, and all the corrections are positive. The unfolding corrections for $\rho^a(r)/\Psi(0.7)$ vary from 0.6 to 1.4, depending on the

Jet p_T	$\|y\| < 0.8$	$0.8 < \|y\| < 1.2$	$1.2 < \|y\| < 2.1$	$2.1 < \|y\| < 2.8$
40-60 GeV	1.007	1.019	1.018	1.050
60-80 GeV	1.011	1.027	1.028	1.048
80-110 GeV	1.012	1.028	1.029	1.045
110-160 GeV	1.014	1.027	1.030	1.041
160-210 GeV	1.014	1.027	1.031	1.035
210-260 GeV	1.013	1.026	1.028	1.030
260-310 GeV	1.013	1.027	1.027	1.023
310-400 GeV	1.011	1.024	1.027	1.020
400-500 GeV	1.011	1.022	1.025	1.016
500-600 GeV	1.008	1.020	1.024	0.999

Table 5.3: Correction factors for the jet shape variable $\Psi(0.3)/\Psi(0.7)$.

p_T, y, and r. Generally, the corrections are smaller for more central jets and are maximal in the first two radial bins.

The corrections for $\rho^a(r)/\Psi(0.7)$ can be explained by a few simple factors. The calorimeter showering and position resolution tend to smear the energy of the jet. As the energy density is highest in the core, the showering tends to move energy into annuli at higher radii. To compensate for this effect, the first bin must be adjusted upwards, and the next several bins downwards. Far from the core of the jet, there is very little energy deposited. In this region, the calorimeter thresholds and noise suppression, as well as the magnetic field in the inner detector (which, in effect, rejects charged particles with p_T <400 MeV) remove some of the energy. Thus, the measured jet energy response must be adjusted upwards where these effects dominate.

5.3.1 Unfolding Bias and Variance

For any unfolding technique there is a bias, owing to the incomplete ability to extract "true" observables (i.e. the correct value of the observable in nature) from reconstructed (smeared) data. For bin-by-bin unfolding, the bias in bin i, b_i is simply the difference between the predicted and true observable [175]:

$$b_i = C_i \times R_i - O_i \tag{5.3}$$

$$b_i = \frac{O_i^{MC}}{R_i^{MC}} \times R_i - O_i$$

Radius	40-60 GeV	60-80 GeV	80-110 GeV	110-160 GeV	160-210 GeV	210-260 GeV
0.0-0.1	1.247	1.258	1.249	1.242	1.228	1.217
0.1-0.2	0.818	0.754	0.719	0.681	0.648	0.622
0.2-0.3	0.816	0.786	0.767	0.742	0.733	0.719
0.3-0.4	0.883	0.854	0.845	0.816	0.806	0.795
0.4-0.5	0.983	0.947	0.925	0.899	0.880	0.864
0.5-0.6	0.984	0.975	0.959	0.949	0.932	0.914
0.6-0.7	1.195	1.173	1.161	1.127	1.108	1.096
0.7-0.8	1.203	1.215	1.213	1.215	1.200	1.189
0.8-0.9	1.236	1.268	1.255	1.254	1.264	1.259
0.9-1.0	1.254	1.295	1.288	1.285	1.290	1.297
1.0-1.1	1.282	1.307	1.327	1.306	1.320	1.323
1.1-1.2	1.285	1.322	1.330	1.330	1.327	1.327

Table 5.4: Correction factors for the jet shape variable $\rho^a(r)/\Psi(0.7)$ for jets in six p_T bins, with $y < 0.8$.

Re-written as a fractional error on O_i, the bias is:

$$\frac{b_i}{O_i} = \frac{O_i^{MC}}{R_i^{MC}} \times \frac{R_i}{O_i} - 1$$

This bias has several key features. Critically, the bias is exactly zero if and only if:

$$\frac{R_i^{MC}}{O_i^{MC}} = \frac{R_i}{O_i} \tag{5.4}$$

That is, the size of the bias does *not* depend on the size of the correction, nor does it depend on the size of the difference between the Monte Carlo simulation and the data. The correction factor must, however, properly account for this difference. For example, a difference in the predicted particle spectra within the jet can produce a large bias. If the Monte Carlo predicts only low-energy hadrons within a jet, the corrections to the jet shapes at the core of the jet will be large (corresponding to the large amount of energy lost in the measurement to particles bending out of the jet or being stopped by the material prior to the calorimeter). If the hadrons within the jets in data are, in fact, high-energy, the correction will incorrectly increase the calorimeter energy measurement in the center of the jet, leading to a biased jet shape measurement.

If we write the results in Monte Carlo simulation as $R_i^{MC} = R_i + \Delta R_i$ and $O_i^{MC} = O_i + \Delta O_i$ (where ΔR_i and ΔO_i are not necessarily small), then the fractional bias can be

written as:

$$\frac{b_i}{O_i} = \frac{O_i + \Delta O_i}{R_i + \Delta R_i} \times \frac{R_i}{O_i} - 1$$

$$\frac{b_i}{O_i} = \frac{1 + \Delta O_i/O_i}{1 + \Delta R_i/R_i} - 1$$

$$\frac{b_i}{O_i} = \frac{\Delta O_i/O_i - \Delta R_i/R_i}{1 + \Delta R_i/R_i}$$

If the detector response reduces the difference between the correct result (i.e. "the truth") and the Monte Carlo ($|\Delta R_i| < |\Delta O_i|$, as is often the case), this bias pulls the unfolded result towards the Monte Carlo. For example, if a positive difference between the correct result and the Monte Carlo ($\Delta O_i > 0$) typically produces a smaller positive difference between data and reconstructed simulation ($|\Delta R_i| < |\Delta O_i|$), then the bias will be negative (i.e. the unfolded result falls between the correct result and the Monte Carlo). On the other hand, if a negative difference between the true result and the Monte Carlo ($\Delta O_i < 0$) typically produces a smaller negative difference between data and reconstructed simulation ($|\Delta R_i| < |\Delta O_i|$), then the bias will be positive (i.e. the unfolded result falls between the correct result and the Monte Carlo).

In order to estimate the size of this bias, in the subsequent sections a variety of samples are considered to test the ability of the unfolding to remove detector effects and correctly predict the result. In each case, the unfolding applied is the one derived using the PYTHIA MC09 Monte Carlo simulation sample. The bias is given by the difference between the jet shapes at hadron-level in the different sample and the shapes unfolded from the reconstructed calorimeter data in the same sample (i.e. Equation 5.3, but using the alternate sample as a replacement for "the truth"). Because of the large number of similar distributions, most of the figures are collected in Appendix 7.2. Only the comparisons demonstrating the largest bias or some particularly interesting feature are included in this chapter.

In attempting to produce an exact unfolding, large changes in response (i.e. significant bin-to-bin migrations between hadron-level and calorimeter-level distributions) may lead to unstable unfolded results. These unstable results appear as bin-to-bin fluctuations in the final result and are called "variance." For smooth hadron-level distributions, and for correction factors near unity, the variance is expected to be small for bin-by-bin unfolding. For the most part, large variance is apparent in the unfolded distribution, and can be evaluated to

a large degree simultaneously with the unfolding bias. If the variance is significantly larger than the systematic uncertainty from the bias, it must be separately taken into account.

5.3.2 Unfolding Bias from Hadronic Physics Models

One of the key effects on calorimeter-based jet shapes is the model used to describe hadronic physics in the calorimetry. Electromagnetic showers are extremely regular, and therefore have a very predictable effect on shower shapes. However, an incorrect description of the showering of hadrons (especially charged pions) could lead to a dramatically altered jet shape.

The ATLAS simulation is based on GEANT4 [153, 154], which provides several collections of physics models (called "physics lists") to users. The ATLAS Collaboration has chosen to use the quark-gluon string precompound model [176, 177, 178, 179] with a Bertini intranuclear cascade [180, 181] as its primary physics list. This list generally provided good agreement with longitudinal and lateral shower shapes in the tile calorimeter test beam [155]. It has several known problems, however, including an unphysical dip in response around 8-10 GeV [182]. It also produces showers for single pions and protons that are too narrow by ∼15%, though this effect should not result in the same deviation in jet shapes, which are dominated by pre-calorimeter showering. These lists are expected to evolve over time as they are tuned, in part with input from the LHC experiments.

In order to test the effect of a different physics list on the jet shapes, di-jet simulation was done with two alternate physics lists. The first, QGSP, simply omitted the Bertini intranuclear cascade. The second, FTFP_BERT, used the Fritiof precompound model [183, 184] with a Bertini intranuclear cascade. This list has a fundamentally different description of nuclear interactions, but the same description of the inner-workings of a nucleus (owing to the Bertini cascade model). These two models are expected to provide some bound on the showering observed in the test beam results [146, 147, 148], in that one is narrower and longer, and the other is wider and shorter than the QGSP_BERT. The result of this comparison can be seen in Figure 5.14 (see also Figure A.1).

The bias in unfolding is exactly the difference between the hadron-level observable in this modified sample and the unfolded calorimeter-level observable. The ratio gives the fractional bias. The systematic uncertainty on the unfolding must encompasses these biases.

The bias in the unfolded shape $\Psi(0.3)/\Psi(0.7)$ is below 2% for all p_T and rapidity bins.

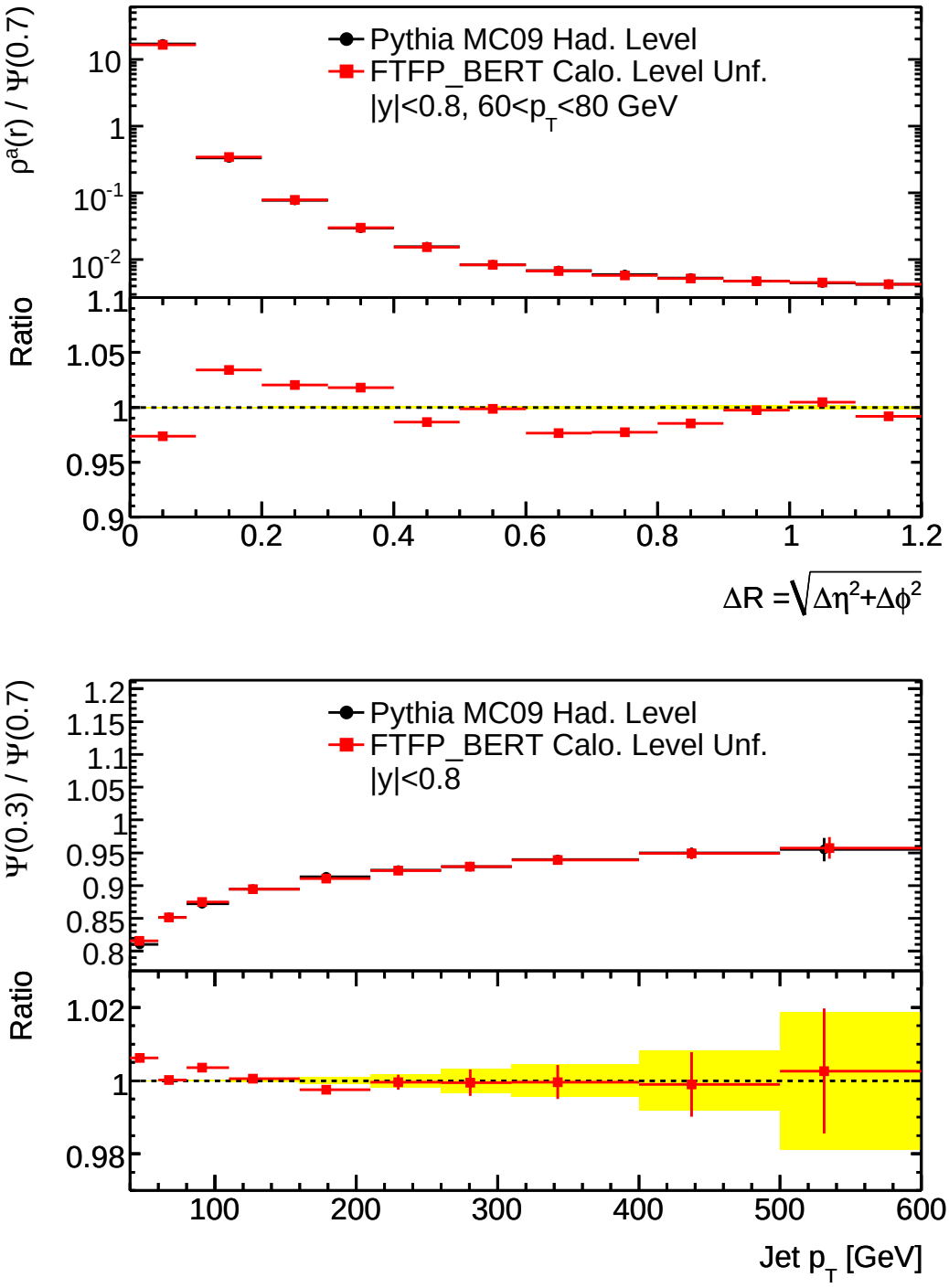


Figure 5.14: Comparison of the jet shapes $\rho^a(r)/\Psi(0.7)$ (top) and $\Psi(0.3)/\Psi(0.7)$ (bottom) with the FTFP_BERT physics list after unfolding using the QGSP_BERT physics list. The difference between the hadron-level distribution and the unfolded calorimeter-level distribution is the unfolding bias in this sample. The ratio is defined as the unfolded calorimeter-level shapes divided by the hadron-level shapes.

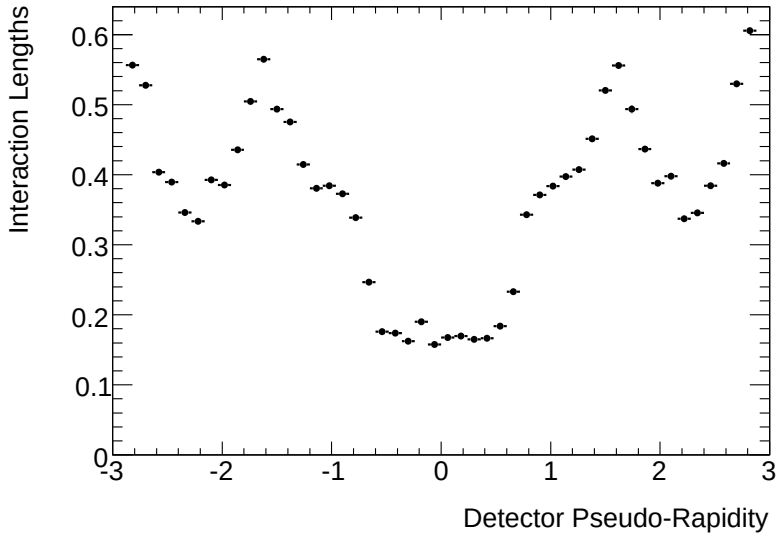


Figure 5.15: The number of interaction lengths before entering an active portion of the calorimeter.

For the shape $\rho^a(r)/\Psi(0.7)$, the bias is up to $\sim$10% and is largest at high radius. The largest effect by far is from QGSP. This physics list produces significantly different lateral single-pion and single-proton shower shapes to those of QGSP_BERT and disagrees with the test beam data. Therefore, we include as the bias from physics lists the bias in the FTFP_BERT sample.

5.3.3 Unfolding Bias from Detector Description

ATLAS has a relatively thick inner detector, and additional passive material in the calorimeter itself, including cryostats and the solenoid magnet. There are several radiation lengths of material between the interaction point and the portion of the calorimeter measuring jets. The number of interaction lengths before entering an active portion of the calorimeter (the presampler layer or the first active layer of electromagnetic calorimetry) is shown in Figure 5.15. As a result, the hadrons produced in the initial collision undergo a significant number of interactions (including nuclear scattering) before reaching the calorimeter. Calorimetric jet shapes are, therefore, sensitive to the details of the geometric description of the material leading up to the calorimeter and the inactive material between layers of the calorimeter.

In order to evaluate the effect of an underestimation of the material, di-jet events were

simulated with a modified detector layout. The LAr calorimeter cryostat was made 20% of a radiation length thicker, both in front of the LAr calorimeter and in front of the tile calorimeter. Additionally, 5% of a radiation length was added between the barrel and the first sampling layer of the calorimeter. The material of the inner detector has been studied in some detail [185, 186]. It is apparent that no more than a few percent of a radiation length of material are missing, and, therefore, the effect of additional material in the inner detector alone has been neglected.

The bias in the unfolded shape $\Psi(0.3)/\Psi(0.7)$ is below $\sim$1% for all p_T and rapidity bins (cf. Figure A.2). For the shape $\rho^a(r)/\Psi(0.7)$, the bias is less than $\sim$10% and is largest at high radius. The effect on $\rho^a(r)/\Psi(0.7)$ is largest at the edge of the jet, where most of the particles are expected to be soft. This is consistent with the expectation that additional material in front of the calorimeter will reject more soft particles.

5.3.4 Unfolding Bias from Generator Tuning

If the generator tuning used for the unfolding differs significantly from nature, the unfolded distribution will be accordingly biased. ATLAS uses a Monte Carlo generator tuning called MC09 [63]. This tune includes modifications to the default Pythia settings for multi-parton interaction models, fragmentation models, and underlying event. It is based on underlying event, minimum bias, and jet analysis data at several center-of-mass energies.

In order to evaluate these differences in previous experiments, frequently portions of the event generator were disabled (the initial state radiation or multi-parton interactions, for example). This provides a gross overestimate of the unfolding bias associated with these models. Worse, the resulting jet shapes are completely unphysical and may not even reflect a "no underlying event" hypothetical reality. All of the models of each generator are tuned *together*, so that without any single model the result is at best unpredictable. The far better approach is to consider several different tunes with some reasonable variation, using different input datasets for tuning if possible, and to test the variation between these tunes. Two independent Pythia tunes were used to evaluate the bias from MC09, one that particularly changes hadronization model parameters and another that particularly changes underlying event model parameters.

In order to evaluate the bias associated with the fragmentation model, a tune provided by the Professor collaboration [62] was used. This tune was made independently of the

MC09 tune, and should, therefore, provide an estimation of the distribution of reasonable fragmentation tunes. In this case, the LEP data were used in the fragmentation tuning in order to minimize the effects of the underlying event on physics results. The differences after unfolding are tiny (cf. Figure A.3), implying a small effect from a change in hadronization. The variation of $\rho^a(r)/\Psi(0.7)$ is consistent with a variance of several percent.

In order to evaluate the bias associated with the underlying event model, a tune called "Perugia0" [64] was used. One of the most difficult pieces of the Monte Carlo to tune is the underlying event. Typically, underlying event includes multi-parton interactions and initial-state radiation. It must also include some description of the complex color-connection between the beam remnant and the scattered partons.

For technical reasons, this tune additionally used a different parton distribution function (PDF). Whereas the standard PYTHIA samples used the MRST 2007 modified LO PDF's (see Chapter 2), Perugia0 used the CTEQ5L leading-order PDF set [38]. As a result, the comparison also tests for differences between quark and gluon content in LO and modified LO (effectively NLO) PDFs. Because of the dependence of jet shapes on quark- and gluon-content of the sample, jet shape distributions are inherently dependent on parton distribution functions.

The bias in the unfolded shape $\Psi(0.3)/\Psi(0.7)$ is below $\sim$1% for all p_T and rapidity bins (cf. Figure A.4). For the shape $\rho^a(r)/\Psi(0.7)$, the bias is less than $\sim$10%. Although this modification to the underlying event produces a different amount of activity far from the jet relative to the PYTHIA MC09 tune, the unfolding factors appear to take it into account.

5.3.5 Unfolding Bias from Transverse Momentum Spectrum

The unfolding method relies on approximate similarity between the data and Monte Carlo used to calculate the unfolding. Because of the jet shapes' dependence on jet kinematics, if there are significant differences in the kinematics of the samples, the jet shapes will inherently not match. In order to evaluate any bias introduced, the p_T of reconstructed jets was increased (decreased) by one standard deviation in the jet energy scale. This increase covers the difference in calibrated jet p_T spectra between data and Monte Carlo simulation seen in Figure 5.4. The unfolded jet shapes in these samples were compared to the hadron-level jet shapes modified in the same way. The difference between the unfolded shapes and the hadron-level shapes corresponds to the bias introduced if the jet p_T spectrum in the

data is different from that of the Monte Carlo.

The bias is the difference between the unfolded and hadron-level distributions for jets with higher p_T and the corresponding difference for jets with lower p_T. The bias is below 5% in both jet shape variables, for all bins of radius, p_T, and rapidity (cf. Figures A.5 and A.6). This test implies that kinematic differences between data and Monte Carlo simulation *alone* are not likely to produce significant variations in the unfolded jet shapes. The jet p_T resolution was found to be consistent between Monte Carlo simulation and data [174], so no further bias is considered from p_T resolution effects.

There is only a few percent difference between jet shapes in the different rapidity regions considered. Thus, any bias from the rapidity spectrum is neglected.

5.3.6 Unfolding Bias from Generator Differences

Different Monte Carlo event generators have different schemes for describing multi-parton interactions and for describing the non-perturbative hadronization after the QCD showers. Pythia and Herwig [56, 57] have fundamentally different descriptions of the hadronization process and may thus produce different jet shapes even when the parton shower is identical. Herwig uses the underlying event model of Jimmy [163]. Herwig++ [58] is a rewritten version of Herwig with some of the same underlying physical assumptions (e.g. angular ordering of parton showers) but with improved implementations of the models for parton showering and hadronization. Herwig++ does not use Jimmy for modeling of the underlying event. In the previous sub-section, differences in kinematics were found to produce a small effect on the unfolded jet shapes. Thus, the kinematic differences between these generators alone is not likely to produce a large bias.

In order to evaluate these differences, jet shapes were compared in di-jet Monte Carlo produced with Pythia, Herwig, and Herwig++. The same requirements were placed on jets built from stable hadrons in all cases, and the same detector simulation was applied to all models. Figure 5.16 shows a comparison of the unfolded jet shapes and hadron-level jet shapes between the generators (see also Figure A.7).

The bias introduced in the Herwig and Herwig++ samples by the unfolding of $\Psi(0.3)/\Psi(0.7)$ is less than 3% and falls steeply with jet p_T in all rapidity bins. The bias introduced in $\rho^a(r)/\Psi(0.7)$ is up to ∼20% in the radial shapes. This is by far the largest bias of any sample. The difference is most dramatic farther from the center of the jet, where the

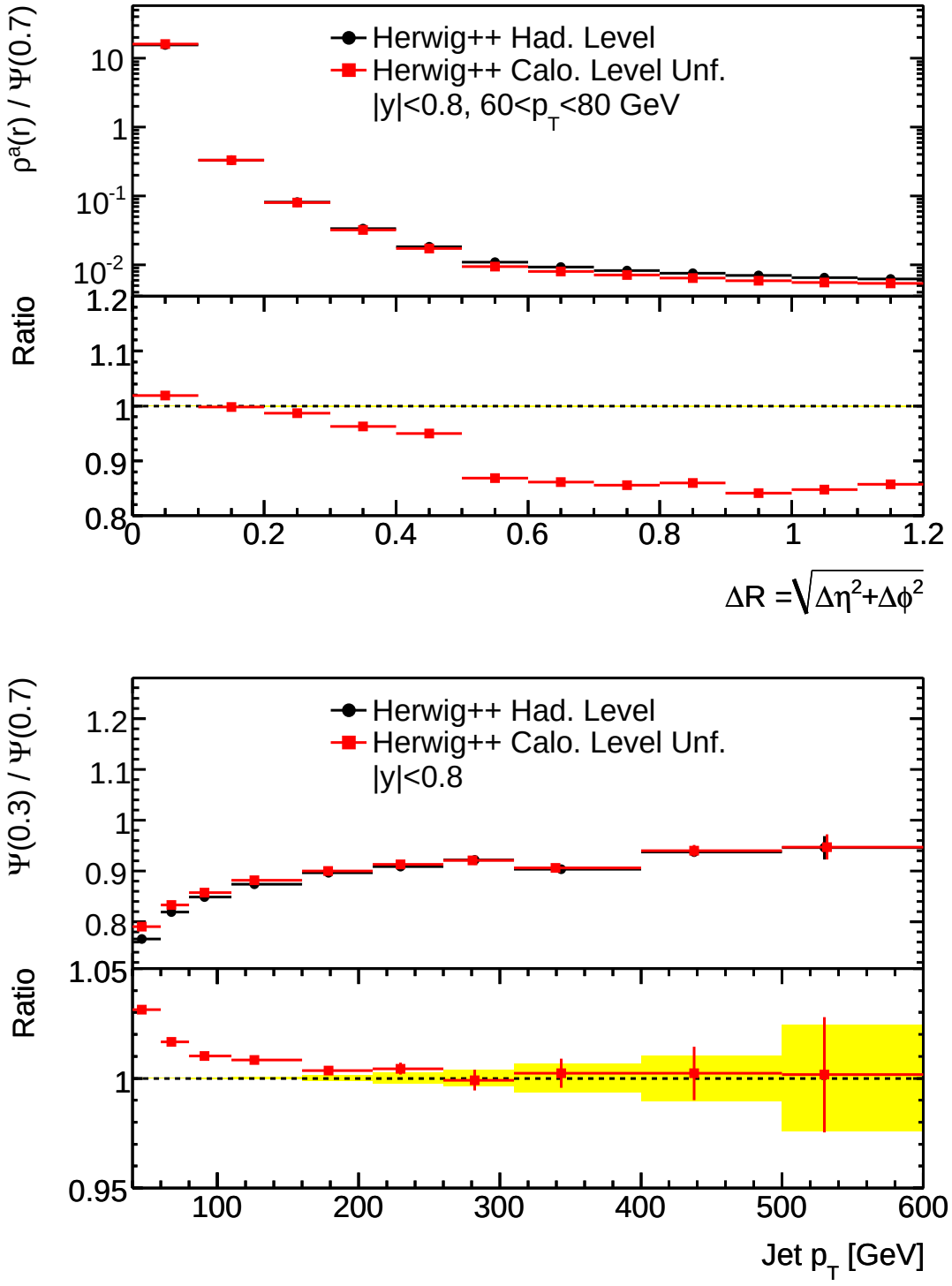


Figure 5.16: The jet shapes $\rho^a(r)/\Psi(0.7)$ (top) and $\Psi(0.3)/\Psi(0.7)$ (bottom) in Herwig++ at the hadron level and after unfolding based on Pythia's MC09 tune. The ratio is defined as the unfolded calorimeter-level shapes divided by the hadron-level shapes.

underlying event and hadronization models make the most significant difference in the jet shapes. The raw difference in $\rho^a(r)/\Psi(0.7)$ between these two generators is ~50% at high radius and low p_T. Neither the PYTHIA nor the HERWIG++ generator tunings included jet shapes as a part of the input data set. Therefore, it is likely that re-tuning including the Tevatron data or the LHC measurements would produce a dramatically better agreement. Although at the Tevatron better agreement was found after generator tuning [96, 97, 98], the ATLAS Collaboration's choice to use the newer p_T ordered shower model in PYTHIA may necessitate considerable re-tuning.

The ALPGEN [42] generator uses the hadronization and underlying event model of HERWIG and JIMMY but generates the matrix-element events differently. Rather than generating 2→2 processes and relying on leading-log parton showers to provide a model for the 2→N processes, ALPGEN includes 2→N at the matrix-element level for N<6. ALPGEN event generation also used a different matching scheme (MLM rather than CKKW, see Chapter 2), a different tune of HERWIG and JIMMY, and different PDFs. The bias is most dramatic at low p_T and low rapidity for $\Psi(0.3)/\Psi(0.7)$, and at low p_T and large radius for $\rho^a(r)/\Psi(0.7)$ (cf. Figure A.8). As the number of nearby jets, and even the number of total jets, is comparable for the two samples, the difference should be due to the underlying event model provided by JIMMY or the difference in kinematics between the two samples (ALPGEN produces somewhat harder and somewhat less central jets). In particular, the bias introduced in the ALPGEN sample is almost identical to the one introduced in the HERWIG sample.

Each of these generators includes slightly different settings for considering a hadron to be "stable." Particular where the contribution from soft particles is important, these differences could have some effect on the hadron-level jet shapes. In order to evaluate any possible effect, jet shapes were constructed using hadrons above several kinematic cut-offs. No effect on the jet shapes were found when removing particles with p_T below 200 MeV, and the effect was on the few percent level when remove hadrons below 500 MeV. Thus, any differences due to this setting should be minimal.

5.3.7 Unfolding Bias from Vertex Position

For events with a single vertex, it is possible to test the quality of the correction for vertex displacement in z. Such displacement will result in a shift in the p_T of the jet, as well as a change in the angle between constituents. It may also result in more or less material

between the interaction point and the calorimeter, affecting the evolution of the jet shape in the detector. In order to test this bias, two comparisons were made, both using Monte Carlo simulation. Data samples, similarly constructed, would include different amounts of background, and, therefore, the comparison would be problematic.

A modified di-jet event sample was simulated using a luminous region centered at (1.5,2,-9) mm in (x, y, z). The reconstructed jet shapes in this sample were unfolded and compared to the hadron-level jet shapes constructed without the luminous region shift. The bias introduced by the unfolding is well below 1% in $\Psi(0.3)/\Psi(0.7)$ and below 5% in $\rho^a(r)/\Psi(0.7)$ (cf. Figure A.9).

Additionally, jet shapes were constructed using jets with reconstructed vertex location $|z|$ <5 mm, $|z|$ <50 mm, $|z|$ >50 mm, and $|z|$ >100 mm. Figure 5.17 shows a comparison of the jet shapes in these different samples, in Monte Carlo simulation. In both shape variables, the difference between the first two samples is negligible, so only $|z|$ <50 mm is shown. The difference between the first three samples is less than 0.5% in $\Psi(0.3)/\Psi(0.7)$ and below ∼10% in $\rho^a(r)/\Psi(0.7)$. The last sample ($|z| > 100$ mm) is removed by the primary vertex cut.

Conservatively, the bias in the unfolding is estimated as the difference between the $|z|$ <50 mm and $|z|$ >50 mm samples, multiplied by the difference in fractional population (<5%). The corresponding bias on $\Psi(0.3)/\Psi(0.7)$ is below 0.1% for all rapidity and p_T bins. On $\rho^a(r)/\Psi(0.7)$, the bias is below 0.5% for all p_T, rapidity, and radius bins. This is almost negligible next to the biases from generator differences.

5.3.8 Unfolding Bias from Differences in Position Resolution

The jet position resolution is one of the largest factors influencing the unfolding corrections. If the position resolution is different between Monte Carlo and data, then the unfolded distributions will be biased. The effect of a difference in position resolution between data and Monte Carlo simulation was tested using hadron-level jet shapes before and after smearing the jet axis.

For each jet, the jet shape was constructed the first time using the reconstructed jet axis. Then, the jet axis was shifted in y and ϕ by a random number generated from a Gaussian distribution with width 0.025. This width is approximately the jet position resolution for jets around 80 GeV and should overestimate the difference between data and Monte Carlo

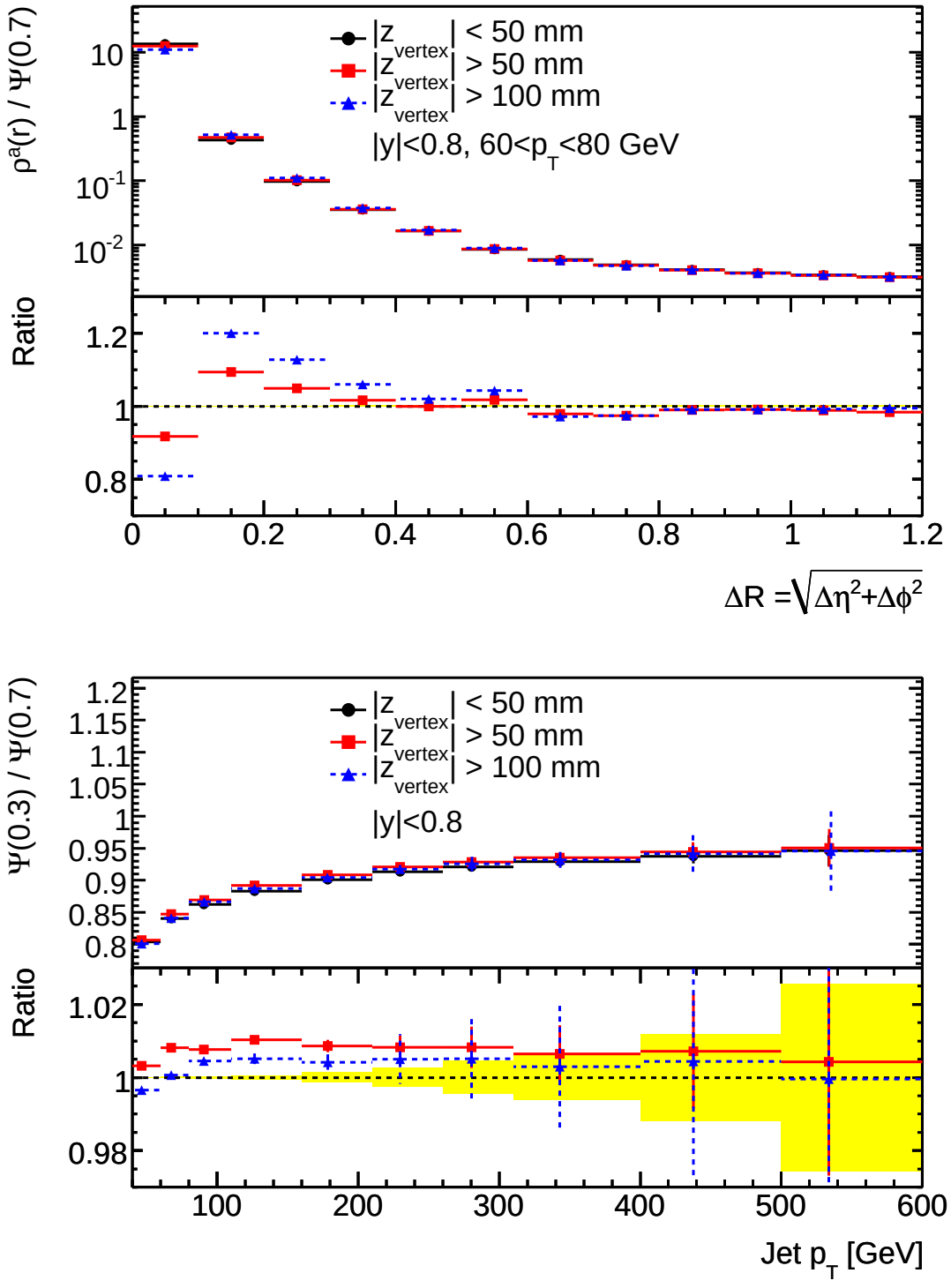


Figure 5.17: Comparison of unfolded reconstructed and hadron-level jet shapes in events with a displaced reconstructed vertex. The ratio is defined with respect to the sample with the most central z distribution.

simulation. A difference in position resolution will typically result from a poor description of noise, calorimeter showering, or underlying event. In each of these cases, the position resolution is worsened because of a difference in energy deposition in the calorimeter. This artificial smearing of the jet axis, therefore, is an extremely pessimistic limit on the effect of deteriorated position resolution. Eventually, the position resolution will be measured in data, but without sufficient statistics it has not yet been possible. From a track-cluster-matching analysis, an upper bound on the difference in ϕ position resolution between data and Monte Carlo simulation was set at 10%, or ∼0.005 radians [187].

The effect of smearing the jet axis on $\Psi(0.3)/\Psi(0.7)$ is tiny (cf. Figure A.10). The jet shape $\rho^a(r)/\Psi(0.7)$ is more affected, particularly in two pairs of bins. The first and second bins shift, where the core of the jet is displaced in the shifted sample. The sixth and seventh bins shift as well, where the edge of the jet is no longer at a local minimum of energy deposition.

5.3.9 Unfolding of Sub-samples

Additionally, it is valuable to check that the unfolding behaves correctly for sub-samples of jets. The unfolding should only take into account distortions due to detector and experimental effects, and therefore should be roughly independent of the underlying physics. In order to evaluate this effect, reconstructed jet shapes after unfolding corrections were compared to jet shapes at the hadron level for various (sub-) leading jet samples. The results of the comparison can be seen in Figures 5.18 and 5.19 for the leading and second-leading jets (see Figures A.11 and A.12 for the third- and fourth-leading jets).

The unfolding derived using the inclusive sample correctly unfolds the sub-leading jet samples to better than 1% in most bins of $\Psi(0.3)/\Psi(0.7)$. The unfolding of $\rho^a(r)/\Psi(0.7)$ is good to ∼10% for the second-leading jet, but for the third- and fourth-leading jets there are significant problems outside the jet cone. Because these jets are not necessarily isolated, there is a fair chance that the third- or fourth-leading jet will be nearby one of the higher p_T jets in the event. This bias in unfolding will be taken as a part of the systematic uncertainty in any case where sub-leading jet samples are considered. As the number of good jets per event is in good agreement between data and Monte Carlo simulation, we do not add any systematic uncertainty on the inclusive sample.

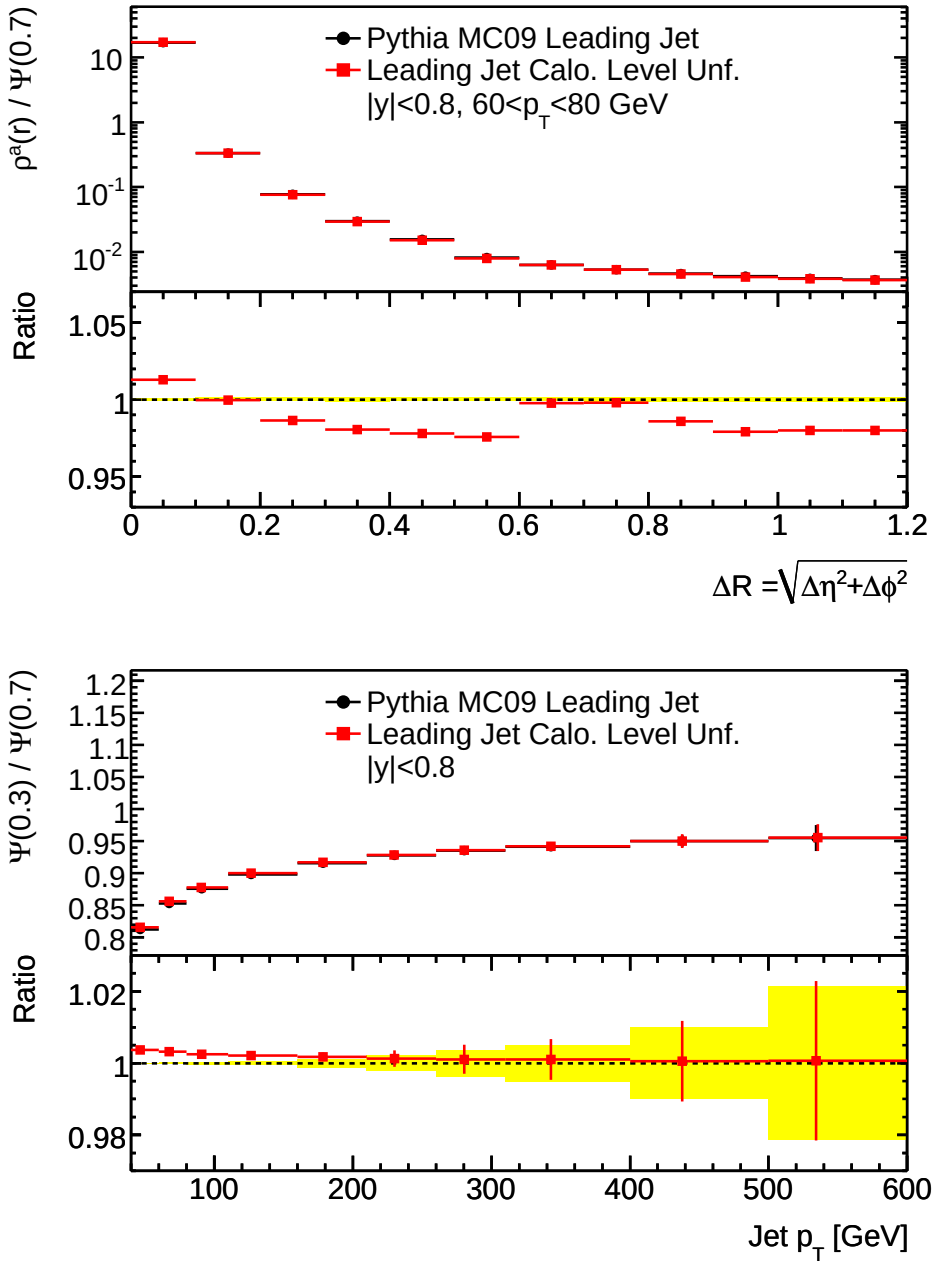


Figure 5.18: Comparison of leading and second-leading jet shapes in Monte Carlo at the hadron level and after detector simulation and unfolding. The unfolding is based on the inclusive sample. The ratio is defined as the calorimeter-level distribution divided by the hadron-level distribution.

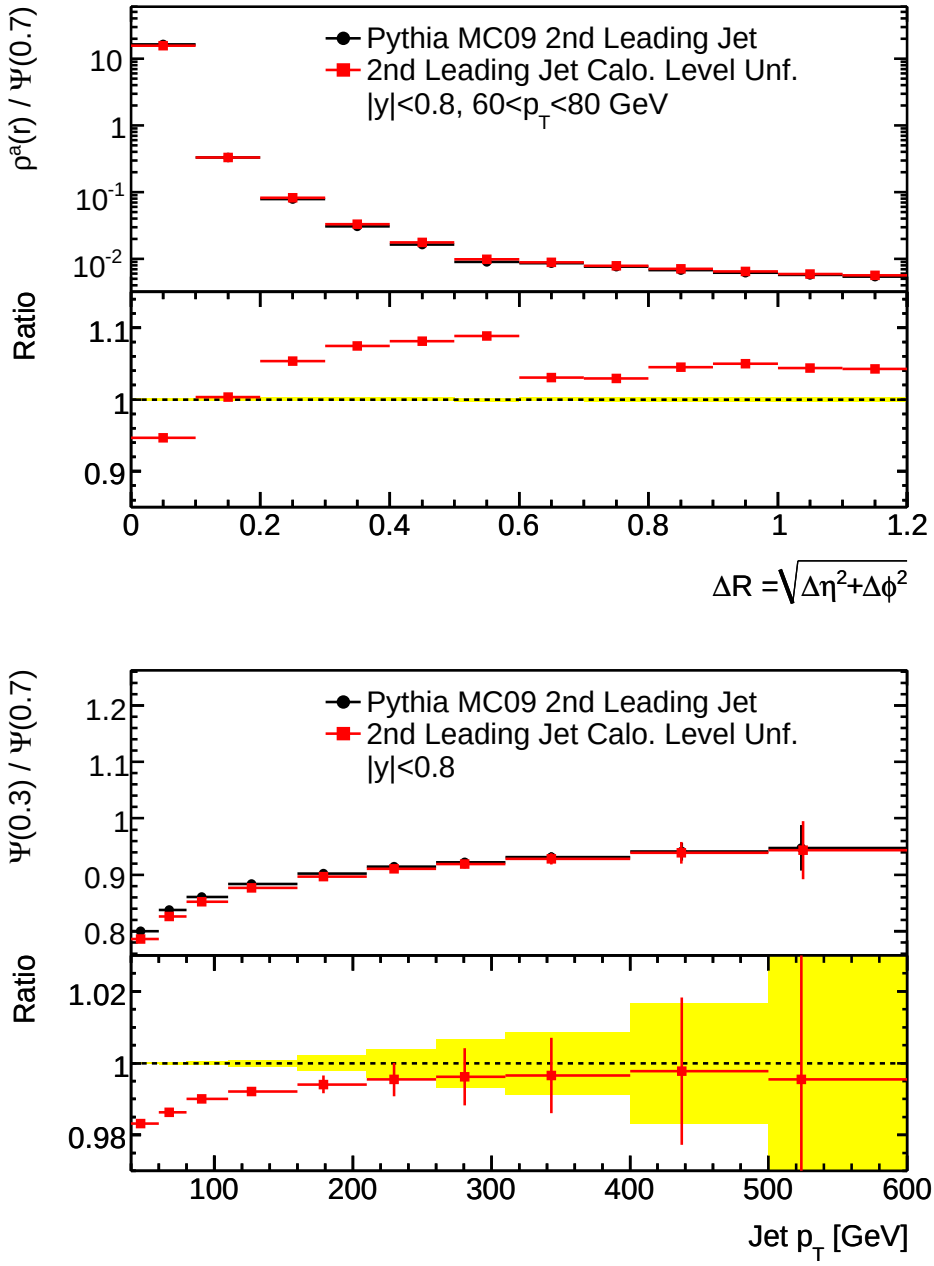


Figure 5.19: Comparison of leading and second-leading jet shapes in Monte Carlo at the hadron level and after detector simulation and unfolding. The unfolding is based on the inclusive sample. The ratio is defined as the calorimeter-level distribution divided by the hadron-level distribution.

5.3.10 Systematic Uncertainty from the Unfolding

Each of the comparisons above provides a measure of the bias in the bin-by-bin unfolding. The biases above are added together in quadrature, and the combination is taken as the systematic uncertainty. In the case of the kinematic variation (up and down in jet p_T), only the worst of the two is taken for a given bin. This combination neglects cases where the bias might be correlated (e.g. increasing the material in the detector and changing the physics list affect the width of the shower in a correlated way). However, it includes some double-counting (e.g. the comparison between generators already includes some differences in tuning and some differences in kinematics).

The systematic uncertainty from unfolding on $\Psi(0.3)/\Psi(0.7)$ ranges from 1-4%, and decreases with increasing jet rapidity and p_T. The systematic uncertainty from unfolding on $\rho^a(r)/\Psi(0.7)$ ranges from 3-25%, increases with jet p_T, and is largest outside the jet. The dominant source of uncertainty in both cases is from the bias in unfolding Herwig++. Once all of the generators have been re-tuned to include measurements from the LHC, this bias may be reduced.

It may be possible to reduce this systematic uncertainty and to better treat possible correlations and double-counting in the measured biases with a more advanced unfolding technique. An example of such a technique will be discussed in Section 7.1.

5.4 Additional Systematic Uncertainties

Here, we consider the three main sources of additional systematic uncertainty: the jet energy scale systematic uncertainty, the uncertainty associated with jet cleaning, and the uncertainty associated with the trigger used for these data.

The ATLAS calorimeter has been well-designed to measure high-p_T objects with good response and linearity. However, as the energy deposited in the calorimeter cell decreases, the magnitude of the error on the energy measurement due to the calibration system and readout electronics of the calorimeter increases. The variation in signal due to *physics* fluctuations (e.g. fluctuations in showering or deposition in the active or inactive portions of the calorimeter) should be well-modeled by the detector simulation, and any uncertainty in these fluctuations are taken into account in the systematic uncertainty on the unfolding. One issue not accounted for in that uncertainty, however, is the quality of the cell calibration

at low energy. Because most of a hadronic jet's energy is captured by the LAr calorimeter, the discussion here will focus on the parameters of the LAr calorimeter calibration.

The uncertainty in the liquid argon calorimeter cell energy calibration is about 20 MeV across the entire dynamic range of the cell, with no apparent bias [143]. The place where this error represents the largest fractional variation in cell response is in the edge of the shower, where very little energy is deposited in any single cell. In that region, however, the annular energy is averaged over many cells due to the large area covered. Since there is no bias in the cell linearity, the error on the measurement of energy collected in all cells around the edge of the shower should be negligible. We therefore do not incorporate a systematic uncertainty for this effect.

5.4.1 Jet Energy Scale Uncertainty

Because of the p_T dependence of jet shapes, any error in the jet energy scale will affect the jet shape distributions. However, because shapes were constructed from EM scale objects, the main impact of jet energy scale uncertainties on jet shapes is in p_T binning. That is, comparing 45 GeV jets in data to 50 GeV jets in Monte Carlo simulation could result in the conclusion that jets in data are wider than those in Monte Carlo.

This effect was evaluated in Monte Carlo by comparing jet shapes binned using the correct jet p_T to jet shapes binned using too-high or too-low p_T. The p_T of the jets was adjusted by one standard deviation in the jet energy scale [170]. The effect of an upward and downward adjustment on the unfolded jet shapes for $\Psi(0.3)/\Psi(0.7)$ and $\rho^a(r)/\Psi(0.7)$ is shown in Figure 5.20 for a single bin of each variable.

The systematic uncertainty on $\Psi(0.3)/\Psi(0.7)$ is below 1.2% for all bins of p_T and y, and falls as p_T rises. The systematic uncertainty on $\rho^a(r)/\Psi(0.7)$ is below 10% for all bins of r, p_T, and y, and falls as p_T rises. This is already almost negligible relative to the unfolding systematic uncertainty and will be further reduced when jet calibrations derived *in situ* are available.

5.4.2 Trigger Biases

The triggers used in this analysis select trigger jets, built from towers of size 0.1×0.1 in $\eta \times \phi$. The individual jet trigger items additionally specify the size of the trigger jet, from 0.4×0.4 to 0.8×0.8 in $\eta \times \phi$. In the case of the L1_J5 trigger item, all trigger jets are 0.4×0.4

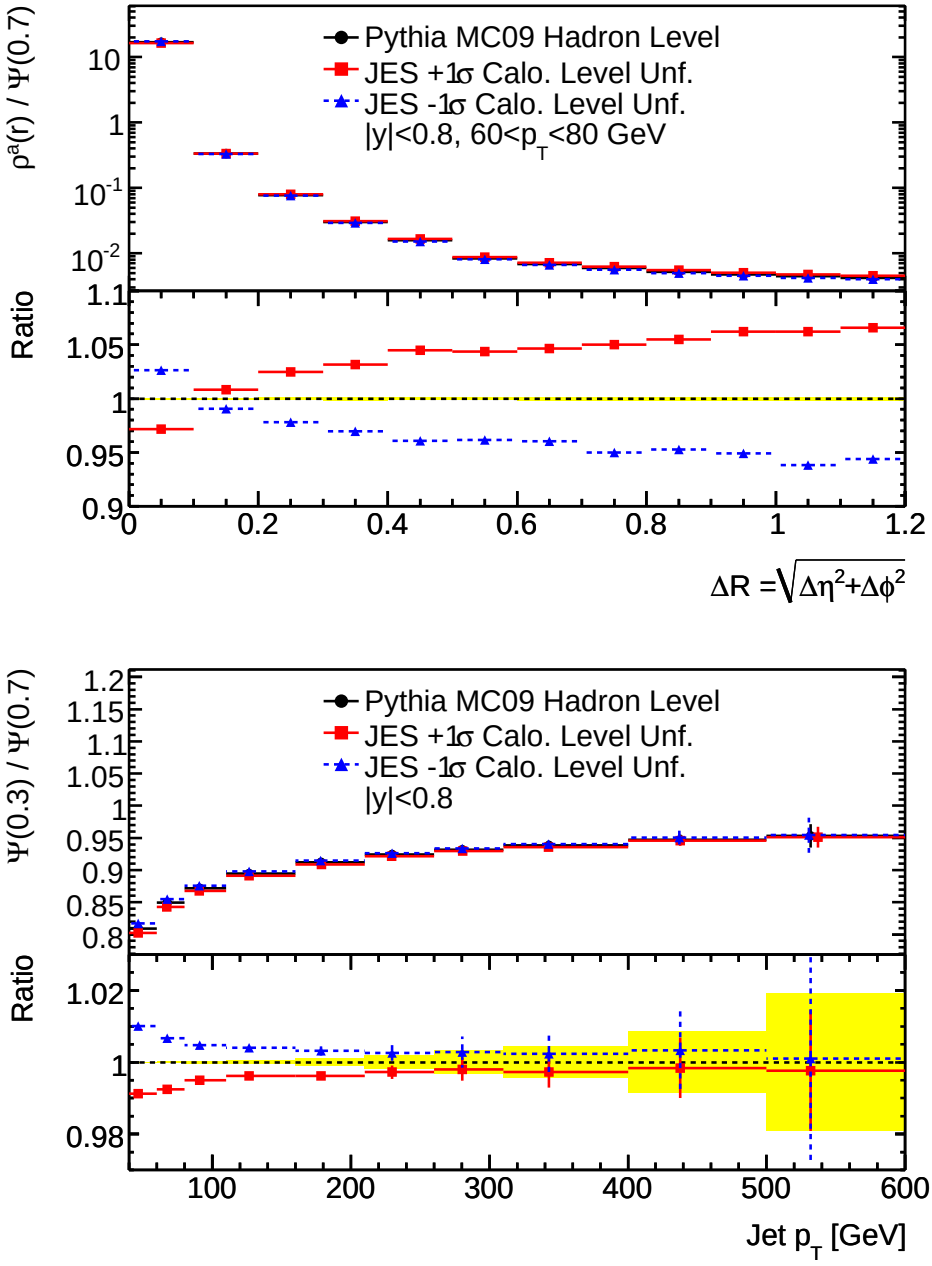


Figure 5.20: Comparison of jet shapes after unfolding using the standard jet energy calibration, a calibration that is higher by one standard deviation of the jet energy scale systematic uncertainty, and a calibration that is lower by one standard deviation. The ratio is defined with respect to the standard calibration.

in $\eta \times \phi$. This different granularity, as well as the significantly different area of the trigger jets, could result in the selection of differently shaped (e.g. more collimated) jets.

The L1_J5 trigger reaches 99% efficiency between 50 and 60 GeV of transverse momentum for offline Anti-k_t jets with D=0.6. A narrow jet below this p_T is more likely to cause a trigger, as a sufficient energy is deposited within the area of a trigger jet. Therefore, for jets causing a trigger below the full efficiency point of this trigger, there is a bias towards narrower jets.

If jets are required to not overlap with a trigger jet in the event, however, an anti-bias could be introduced in low p_T jets (i.e. a rejection of all narrow jets, and selection of only wide jets). For example, consider a low p_T di-jet event in which the two jets were well-balanced in p_T, but in which one jet was more collimated than the other. The more-collimated jet might produce a trigger jet while the wider jet might not. The requirement that a jet be "away" from a trigger jet would therefore select wider jets and induce a bias in the low p_T region.

In order to test these biases, three level one triggers were considered. The two level-one jet triggers, L1_J5 and L1_J15, were the triggers used for this analysis and were tested. Additionally, triggers can be provided by two sets of "minimum bias trigger scintillators" on either side of the detector. These scintillators have approximately the same acceptance as the end-cap calorimeter ($2.82 < |\eta| < 3.84$), and registered hits for any charged particle passing through one of their chambers. A trigger requiring only one hit on either side of the detector, L1_MBTS_1, was also tested for bias in jet shapes.

Figure 5.21 shows the turn on of the three level one triggers as a function of jet p_T and the bias in $\Psi(0.3)/\Psi(0.7)$ introduced by these trigger items in Monte Carlo. The L1_J5 trigger was found to be slightly more efficient in data than in Monte Carlo, where it reaches 95% efficiency around 40 GeV and 99% efficiency around 60 GeV. There is no significant bias in $\Psi(0.3)/\Psi(0.7)$ even from 40-60 GeV. The bias in $\rho^a(r)/\Psi(0.7)$ is below ~2% for jets $40 < p_T < 60$ GeV, as anticipated from the efficiency of the trigger in that p_T bin (cf. Figure A.13). Above 60 GeV, the bias from the L1_J5 jet trigger is negligible. A similar effect can be seen in the L1_J15 jet trigger, which turns on above 80 GeV. There was particularly low efficiency in this trigger in the crack regions ($0.8 < |y| < 2.1$). Therefore, the trigger was not used in any rapidity region for jets below 110 GeV, at which p_T it was more than 99% efficient in *all* rapidity regions.

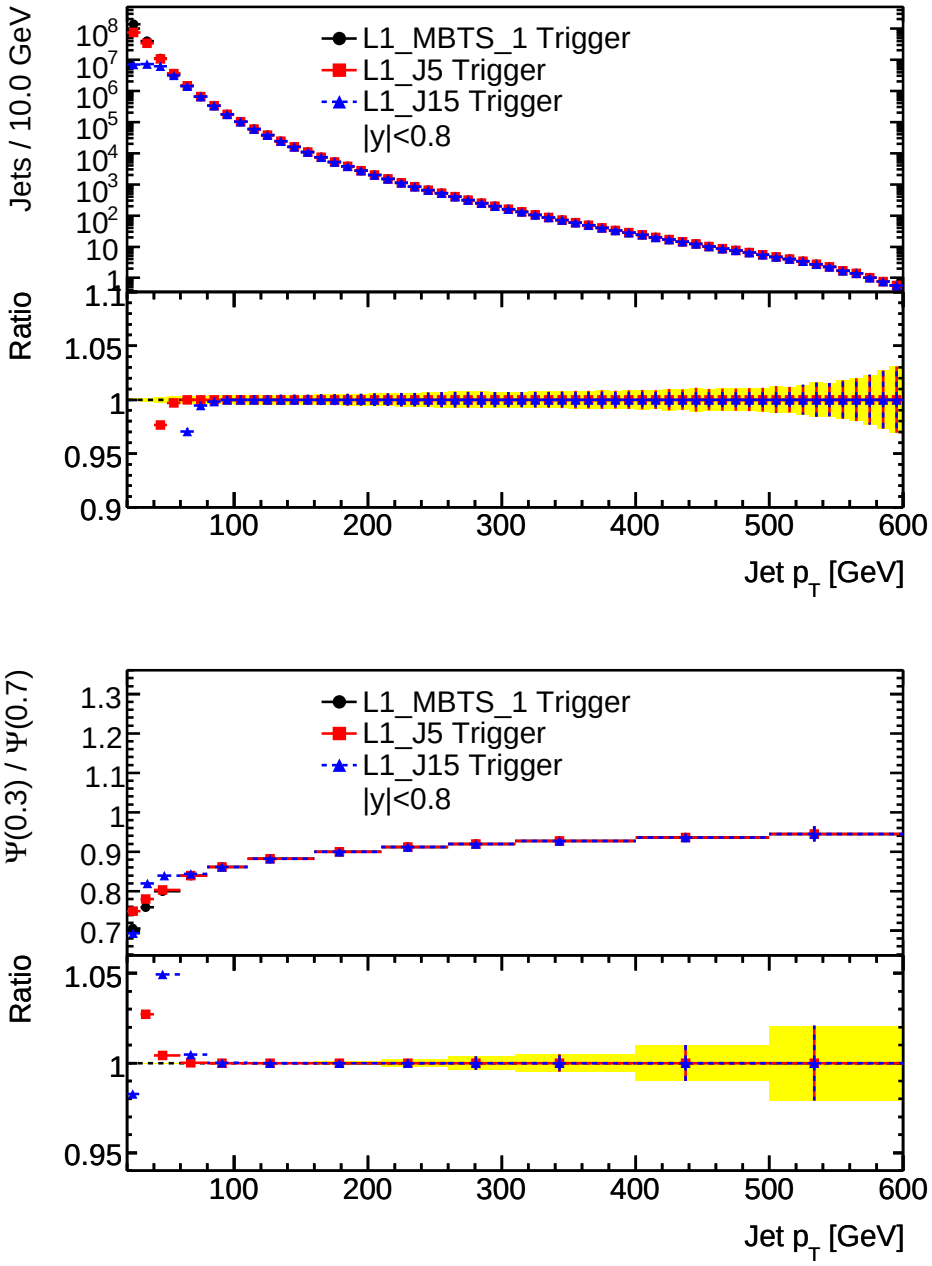


Figure 5.21: Top, the jet p_T spectrum for three triggers of interest in the Monte Carlo simulation. The L1_J5 jet trigger reaches 99% efficiency just below 60 GeV, and the L1_J15 jet trigger reaches 99% efficiency near 80 GeV. Bottom, $\Psi(0.3)/\Psi(0.7)$ as a function of jet p_T for jets in $y < 0.8$. The ratio is defined with respect to the L1_MBTS_1 trigger. The shape is not significantly affected above 40 GeV.

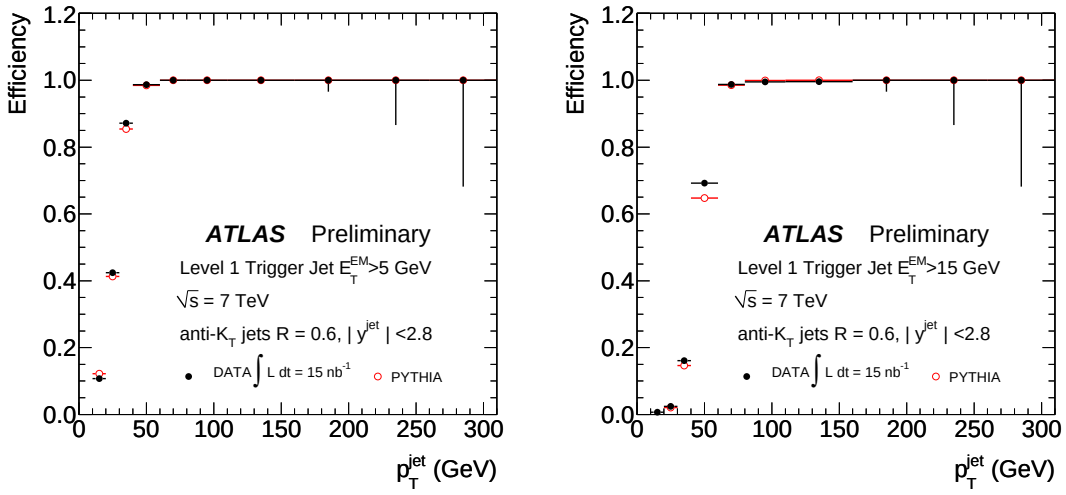


Figure 5.22: The L1_J5 (left) and L1_J15 (right) trigger turn-on curve in data and in Monte Carlo simulation.

These biases should be removed by the unfolding corrections, provided the trigger jet response is the same in data and Monte Carlo simulation. The trigger turn-on curves agree well between data and Monte Carlo simulation, as shown in Figure 5.22 [166]. However, because there is insufficient data to be certain all biases are removed, we conservatively include half of the bias as a systematic uncertainty.

To cross-check these conclusions, the same study was done in data. Data from a single run (i.e. with lower statistics than in the main study) with the minimum bias trigger scintillator trigger were used to construct jet shapes for jets with 20< p_T <60 GeV. These shapes were compared to those from the calorimeter triggers. The results of the comparison are consistent with the results from the Monte Carlo simulation (cf. Figure A.14).

In order to further examine any possible bias in data, four samples of jets were constructed. One sample included all jets in events firing the L1_J5 trigger. A second sample included those jets that did not overlap with a trigger jet in the event (i.e. jets with ΔR_{max}(trigger jet,jet) $\equiv \Delta R^{jt}_{max} > 1.4$). In events with two back-to-back trigger jets, this criterion implies that all jets in the event were accepted. A third sample included all jets which narrowly overlapped with a trigger jet ($\Delta R^{jt}_{min} < 0.2$). This criterion selects jets with an energetic maximum close to the center of the jet. For the purpose of comparing jet shapes at low p_T only, an additional sample was examined, which included all jets which were distinct from a trigger jet ($\Delta R^{jt}_{min} > 1.4$). Because the trigger was highly efficient

above 60 GeV, this sample ran out of statistics in both the data and the Monte Carlo above 80 GeV.

The results of the comparison are shown in Figure 5.23 (see also Figure A.15). As expected, jets that do not trigger are considerably wider than the typical jet in the event. Jets that are close to a trigger jet are a few percent narrower, as anticipated. The Monte Carlo simulation reproduces the differences observed in the data.

5.4.3 Jet Cleaning

The high vertex finding efficiency and low fake rate afford an additional opportunity to check the quality of the jet cleaning cuts used in the data. Jet shapes were compared in events with one good vertex, two vertices, and zero vertices. The rate of "pile-up" events (events with more than one real proton-proton interaction) varied from $\sim 10^{-4}$ for the first runs up to more than 50% for some late runs, and, therefore, this type of event should dominate over events with a fake vertex in the two-vertex sample. Additional proton-proton collisions are unlikely to create high-energy jets, but they may affect the energy of the jets from the primary vertex. In particular, there may be some contribution to the jet shapes in the edges of the jet where little energy is normally deposited. Also, many nearby soft particles from the pile-up events may be combined by the jet algorithm into a single "high energy" jet. From a comparison of the one- and two-vertex events, it is possible to test the evolution of jet shapes under pile-up. The fraction of background events (beam halo, beam gas, and cosmic rays) is higher in the zero-vertex sample than in the one-vertex sample. From a comparison of the zero- and one-vertex events, it is possible to test the effect of the jet cleaning cuts.

In order to evaluate the jet cleaning, jet shapes were compared in data with zero, one, and two reconstructed vertices. The numbers of jets in the data in these samples is shown in Table 5.5. About 6% of the data had more than two reconstructed vertices. All reconstructed vertices were required to have more than four good tracks. The events with zero reconstructed vertices may be background events or may be due to inefficiencies in the vertex reconstruction.

For the 2010 run, the vertex reconstruction efficiency was over 99.9%, and the fake rate was around 10^{-3} [140, 141]. Of course, these fake rates and efficiencies are highly dependent on the event topology. The inefficiency decreases with the number of tracks

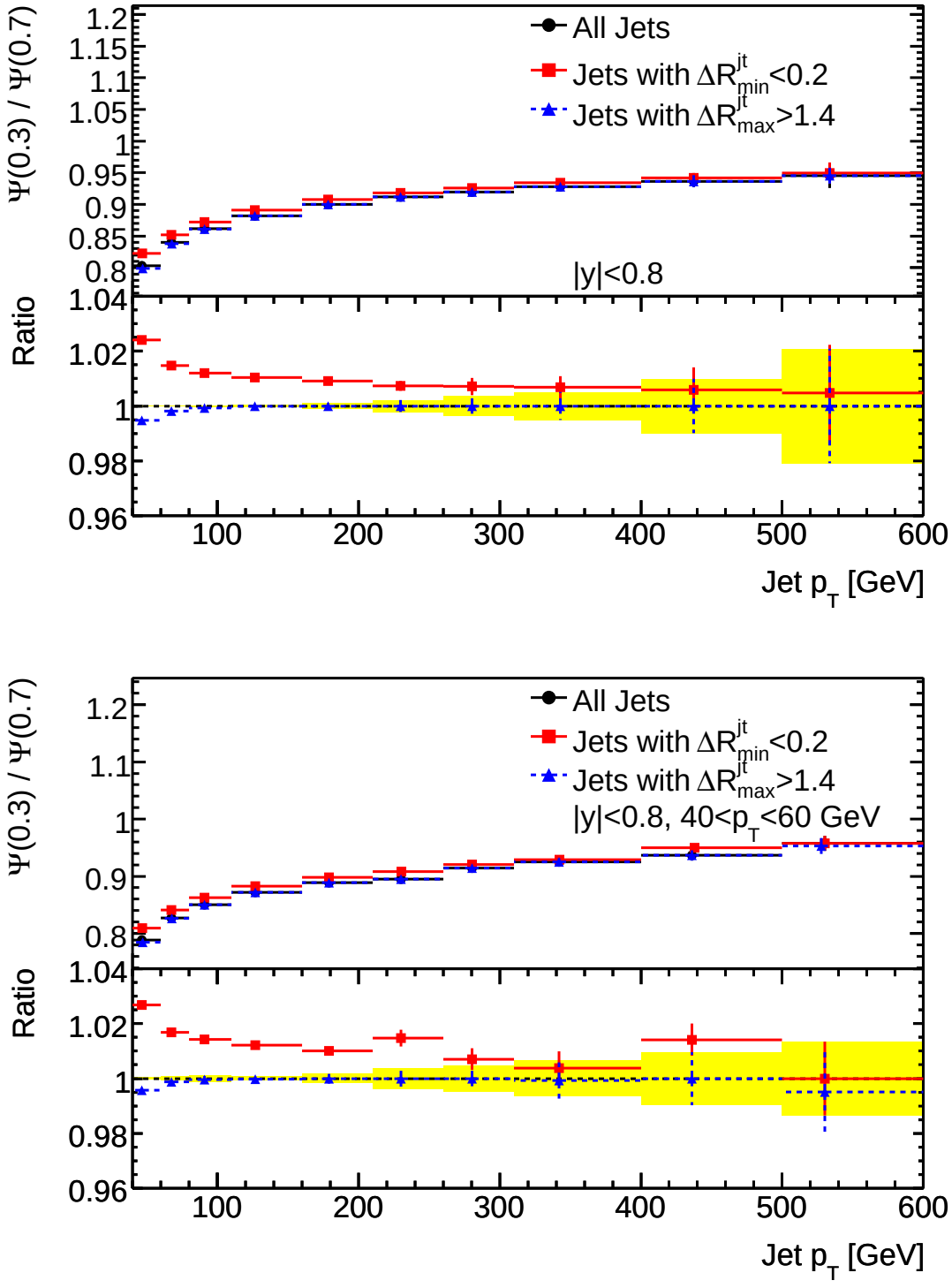


Figure 5.23: Comparison of jet shapes in jets which overlapped in various ways with a trigger jet in Monte Carlo (top) and data (bottom). The ratio is defined with respect to the inclusive sample.

Vertices	Jets after cuts
0	4498
1	851190
2	193113
3	42092
4+	9532

Table 5.5: Numbers of jets and events for data samples with zero, one, two, or three or more reconstructed primary vertices.

in the event, but the fake rate increases. In this case "fakes" are really single vertices that are split into two by the vertex reconstruction software. The ideal efficiency and fake rate for events passing all other event selection criteria can be seen in Monte Carlo simulation in Figure 5.2. Since Monte Carlo simulation only has events with one proton-proton collision, the fraction of events in the Monte Carlo with zero reconstructed vertices provides information on the vertex finding efficiency, and the fraction of events in the Monte Carlo with two reconstructed vertices corresponds to the vertex finding fake rate.

The results for the comparison of $\Psi(0.3)/\Psi(0.7)$ and $\rho^a(r)/\Psi(0.7)$ are shown in Figure 5.24 for a single bin of each variable. As anticipated, the jet shapes in events with two vertices are broader than those in events with exactly one vertex. The difference is most noticeable outside of the jet, where the pile-up events lead to higher background activity throughout the detector. Particularly at low p_T, the jets in events with zero reconstructed vertices are narrower than jets in the single-vertex sample. There is an enhancement of detector noise in this sample, which creates narrower jets than those from collision events. The p_T spectra in zero- and two-vertex events are similar to that of the single-vertex event sample, implying good control of jets from background events. Although the rate is almost 20%, the vertex finding efficiency is sufficiently high that we neglect this uncertainty.

Because of the increasing number of proton-proton collisions per bunch crossing, requiring exactly one good reconstructed vertex may be too restrictive for future jet shape studies. Instead, it may be necessary to correct the jet shapes in some way for the energy deposited by the additional proton-proton collisions in the event. A method for doing so is explored in Section 5.6.5.

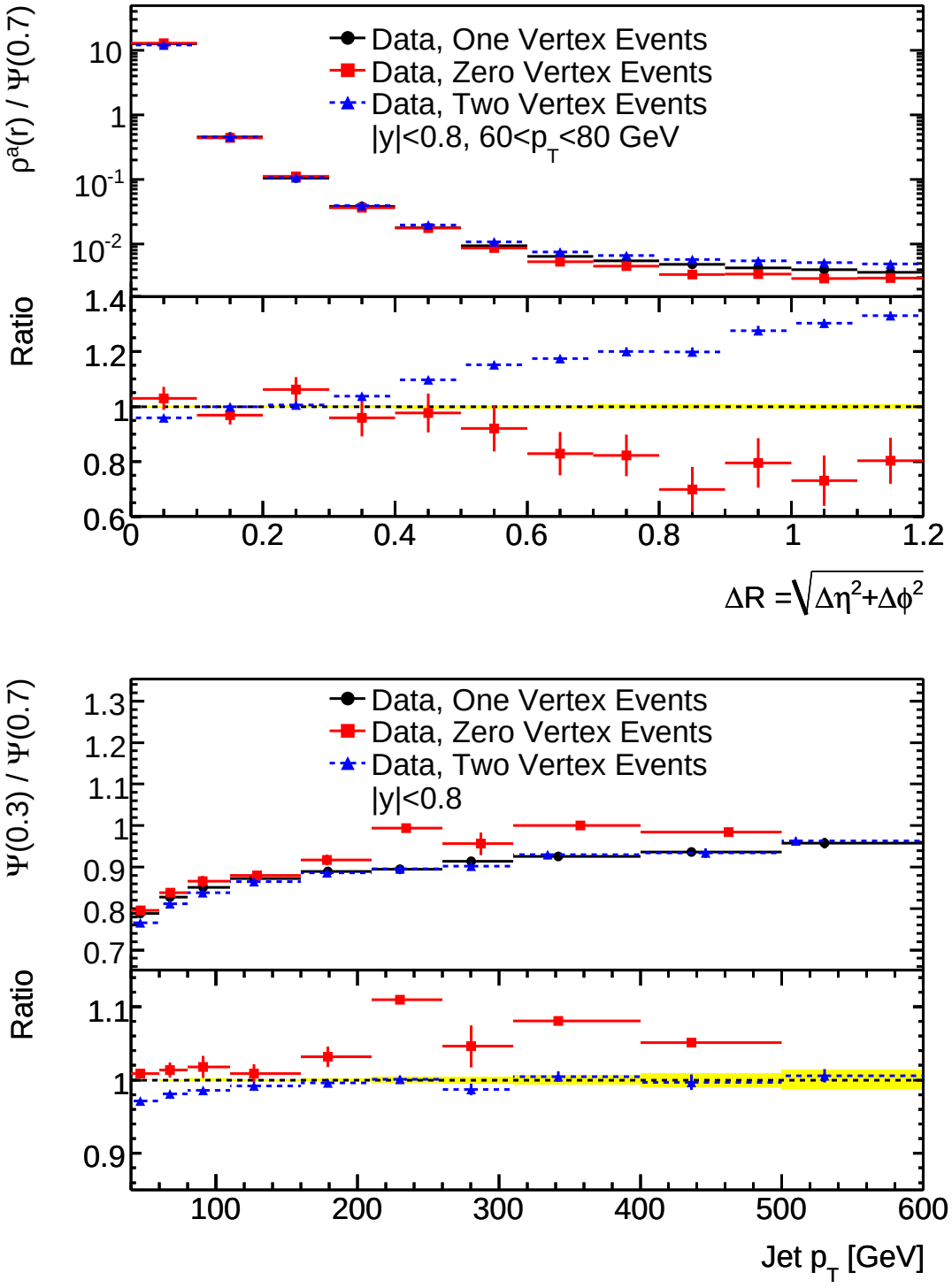


Figure 5.24: Comparison of jet shapes in events with varying numbers of reconstructed primary vertices. The ratio is defined with respect to the single-vertex sample. Two-vertex events are markedly wider than single-vertex events.

5.5 Combination of Systematic Uncertainties

The systematic uncertainties are combined in quadrature for each bin of $\rho^a(r)/\Psi(0.7)$ and $\Psi(0.3)/\Psi(0.7)$. There should be minimal correlation between the various sources of uncertainty. As treating the uncertainty asymmetrically results in a less than 1% difference in the total uncertainty on $\rho^a(r)/\Psi(0.7)$ and a less than 0.01% difference in the total uncertainty on $\Psi(0.3)/\Psi(0.7)$, any asymmetry in the uncertainties is neglected. The final systematic uncertainties are given in Tables 5.6 and 5.7. The total systematic uncertainty on $\Psi(0.3)/\Psi(0.7)$ is below 4% and decreases with jet rapidity and p_T. The total systematic uncertainty on $\rho^a(r)/\Psi(0.7)$ is below ~20% and increases with jet p_T. It is largest outside the jet.

Jet p_T	$\lvert y\rvert < 0.8$	$0.8 < \lvert y\rvert < 1.2$	$1.2 < \lvert y\rvert < 2.1$	$2.1 < \lvert y\rvert < 2.8$
40-60 GeV	3.62	2.99	2.28	1.82
60-80 GeV	2.13	1.84	1.67	1.38
80-110 GeV	1.45	1.91	1.22	1.22
110-160 GeV	1.07	1.19	0.91	1.09
160-210 GeV	0.85	0.74	1.01	0.85
210-260 GeV	0.80	1.03	0.78	0.94
260-310 GeV	0.86	0.90	0.77	1.14
310-400 GeV	0.59	0.59	0.71	1.02
400-500 GeV	0.61	0.52	0.70	1.12
500-600 GeV	0.50	0.68	0.81	4.00

Table 5.6: Systematic uncertainty in % on $\Psi(0.3)/\Psi(0.7)$.

The dominant systematic uncertainty is from the unfolding. Without better tuned Monte Carlo simulation, it is unlikely that this systematic uncertainty can be reduced significantly. Of course, it is also possible that the underlying models of the generators cannot be consistently tuned to all the available data. The current differences between "reasonably tuned" models is up to 50% on $\rho^a(r)/\Psi(0.7)$ in some bins. A more advanced unfolding technique may be able to better account for some of these differences and will be discussed in Chapter 7.

Radius	40-60 GeV	60-80 GeV	80-110 GeV	110-160 GeV	160-210 GeV	210-260 GeV
0.0-0.1	9.0	7.5	5.4	5.1	5.1	4.8
0.1-0.2	9.5	11.5	11.6	13.7	14.3	16.6
0.2-0.3	5.1	7.8	6.9	8.1	8.0	9.0
0.3-0.4	9.8	10.1	6.9	6.6	8.0	9.3
0.4-0.5	11.3	10.3	9.0	7.8	9.3	9.3
0.5-0.6	18.8	16.5	13.7	14.3	9.2	10.0
0.6-0.7	18.4	18.2	13.3	15.1	14.9	26.6
0.7-0.8	21.6	17.9	15.3	13.3	12.0	14.9
0.8-0.9	20.8	17.8	17.0	15.7	12.4	13.7
0.9-1.0	22.2	19.5	19.3	16.1	16.4	13.5
1.0-1.1	22.1	19.7	15.2	19.6	14.2	18.8
1.1-1.2	22.6	18.7	17.2	18.6	17.6	14.6

Table 5.7: The systematic uncertainty in % on $\rho^a(r)/\Psi(0.7)$ for jets with $y < 0.8$.

5.6 Additional Studies and Cross-Checks

Several additional studies were performed that provide cross-checks and further intuition about the physical interpretation of the jet shapes. They do not provide a systematic uncertainty *per se*. Rather, the conclusions of these studies provide confidence that the detector is behaving as expected and that the choices made in this analysis do not produce significant biases.

The calibration of jet constituents was not a part of jet shape studies at any previous collider but is possible with the complex calibration schemes of ATLAS. For example, individual topological clusters of energy can be identified as "hadronic" or "electromagnetic" and calibrated accordingly. The jets, and jet shapes, could then be built from these calibrated topological clusters. These more complex calibrations, however, have not yet been validated using data. As their performance *in situ* is not understood, this study is left as future work.

5.6.1 Geometric Distribution of Constituents

In the calorimeter, jet shapes are evaluated using towers. These towers have finite granularity, and all energy for a single tower is associated with one position (the position of the center of the tower using the Snowmass summation of the tower's cells). No adjustment is

made to these shapes to account for the fraction of the tower which falls inside or outside the area being considered. Stable hadrons, on the other hand, have no pre-defined granularity. The difference is taken into account by the unfolding procedure, but it is possible to explicitly examine the size of this particular component of the correction.

For this test, hadrons were collected into a grid with finite granularity. To approximate the best- and worst-cases for the central calorimeter granularity, grids with characteristic sizes of 0.025 units and 0.1 units in y and ϕ were considered. There is some additional smearing in the calorimeter due to particle showering. As this smearing is on the order of 0.1 units in y and ϕ, it is not likely to dramatically exceed the worst case grid considered here.

The difference for the jet shape $\Psi(0.3)/\Psi(0.7)$ is well below 1% (cf. Figure A.16). The difference for the jet shape $\rho^a(r)/\Psi(0.7)$ makes the reason for this clear. The most significant changes are in the first two bins and the sixth and seventh bins. In the core of the jet, small changes in radius can dramatically change the jet shape. The most likely effect is to broaden the shape of the jet, just as is observed. The increase in the sixth bin and decrease in the seventh bin is consistent with the geometric expectation. In the case of this geometric tower binning, when ever at least ∼40% of a "tower" is contained inside the jet, the entire energy of that tower is included. Near the edge of the jet, where the Anti-k_t jet algorithm places a minimum of energy, this results in the most dramatic change to the measured jet shapes.

5.6.2 Constituent Selection

Jet shapes may be built with different constituents, and jet shapes may be constructed using a variety of constituents as well. In each case, the shapes must still be unfolded to the hadron level. As the dominant systematic uncertainties are on the unfolding itself, rather than the calorimeter noise description, these different starting points for the analysis are unlikely to affect the conclusions. No single answer is "right," but the different shapes might produce subtly different systematic effects, particularly where energy deposition is small and noise suppression is important.

One apparent difference between the data and the Monte Carlo simulation is in the number of cells included in a jet [171]. This difference may come from more cells being slightly above threshold and is likely related to the width of jets measured in the detector.

One possible way to test this difference is to modify the standard noise suppression scheme of ATLAS (see Section 5.2) to omit the final layer of cells around each cluster. This is identical to the "T42" noise suppression scheme used at the Tevatron experiments.

Four alternate jet shape constructions were tested for this study, in order to provide an brief survey of the possibilities. With the standard jets used in this analysis (jets constructed from towers with standard topological noise suppression), jet shapes were built using towers with "T42" noise suppression and with topological clusters. Additionally, jets and jet shapes were built using only "T42" towers or topological clusters.

Topological clusters collect the energy of several towers into a single object. At the core of the jet, this means that the central few towers will be collected, and the jet will appear narrower. This is exactly what is observed in the comparison (cf. Figure A.17). Using topological clusters, the same shapes are constructed using either jets built from topological clusters or jets built from towers. The exception is near the edge of the jet, where the jet finding algorithm has behaved somewhat differently and placed the edge of the jet differently for the two sets of constituents. The "T42" noise suppression scheme gives somewhat narrower jet shapes, particularly far from the jet axis, since it includes fewer cells far from the core of the jet. As the jet p_T increases and the jet becomes more collimated, this difference decreases significantly.

5.6.3 Jet Overlap

For this analysis, no jet overlap requirements were made, in order to better compare to previous studies. It is possible that requiring a jet not overlap with any other jet may introduce a bias in the jet shape by removing the possibility of soft radiation near the jet. However, no requirement may also bias low p_T jet shapes by counting the soft radiation near a hard jet as an independent soft jet. Depending upon the treatment of nearby jets, isolation criteria may produce large changes to jet shapes near the edges of a jet. In the case of two nearby jets, some energy between them must be divided. Typically, the energy is assigned to the closest jet. If jet shapes are constructed using this assignment, however, they will create a narrow shape for each jet. If, on the other hand, jet shapes are constructed using a purely geometric criterion, they will create a wide shape for each jet. The difference exists at both the experimental (calorimeter) and phenomenological (hadron) levels, so provided it is treated consistently the jet shape comparisons will still be valid.

In order to examine this effect, three different "overlap removal" schemes were tested. Unfolding corrections are not applied. Instead, the results for calorimeter-level shapes from Monte Carlo simulation and data are presented uncorrected. In all cases, overlap was defined using the opening angle between two jets passing all cuts (including p_T >40 GeV), ΔR(jet,jet) $\equiv \Delta R_{jj} = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

First, jet shapes were reevaluated for only those jets that were far from other jets, ΔR_{jj} >2.4. In both the Monte Carlo simulation and the data, the difference is most prominent at large radius, where it reaches 10-15% (cf. Figure A.18). Inside the jet, the jet shapes are almost identical. In order to display the difference more clearly, the jet shape $\Psi(1.0)/\Psi(0.7)$ is used. One can see that, if this isolation criterion is used, the out-of-cone corrections for a jet may be different by several percent. In particular, once the jet energy scale systematic uncertainty has been reduced to the several percent level, this difference will need to be taken into account for any exclusive (e.g. exactly one, or exactly three) jet analysis.

Next, jet shapes were reevaluated for only those jets that did not share a part of their "jet cone" with another jet, ΔR_{jj} >1.4. The difference is almost identical to that of the previous comparison (cf. Figure A.19), indicating that the key effect on jet shapes comes when jets are quite close by (ΔR_{jj} <1.4). In both cases, the difference is most important for low p_T jets, simply because high p_T jets are unlikely to have a fractionally large contribution to their momentum from a nearby high p_T jet. Also, when two Anti-k_t jets are nearby one-another, algorithm preserves the cone-shape of the higher p_T jet by cutting away part of the lower p_T jet.

Finally, jet shapes were evaluated for those jets that were far from (ΔR_{jj} >2.4) any jet of higher p_T. This modification allows the possibility of soft radiation outside the jet cone but includes only the main jet and not the radiation in evaluating the shapes. The result of this comparison is shown in Figure 5.25. Again, the difference is most prominent at high radius and low p_T. It is significantly smaller than the previous two differences, however, indicating the importance of distinguishing between nearby high p_T jets and nearby soft jets. Any isolation criterion applied at the experimental level will suffer from this effect, because the calorimeter is unable to resolve jets below a certain p_T.

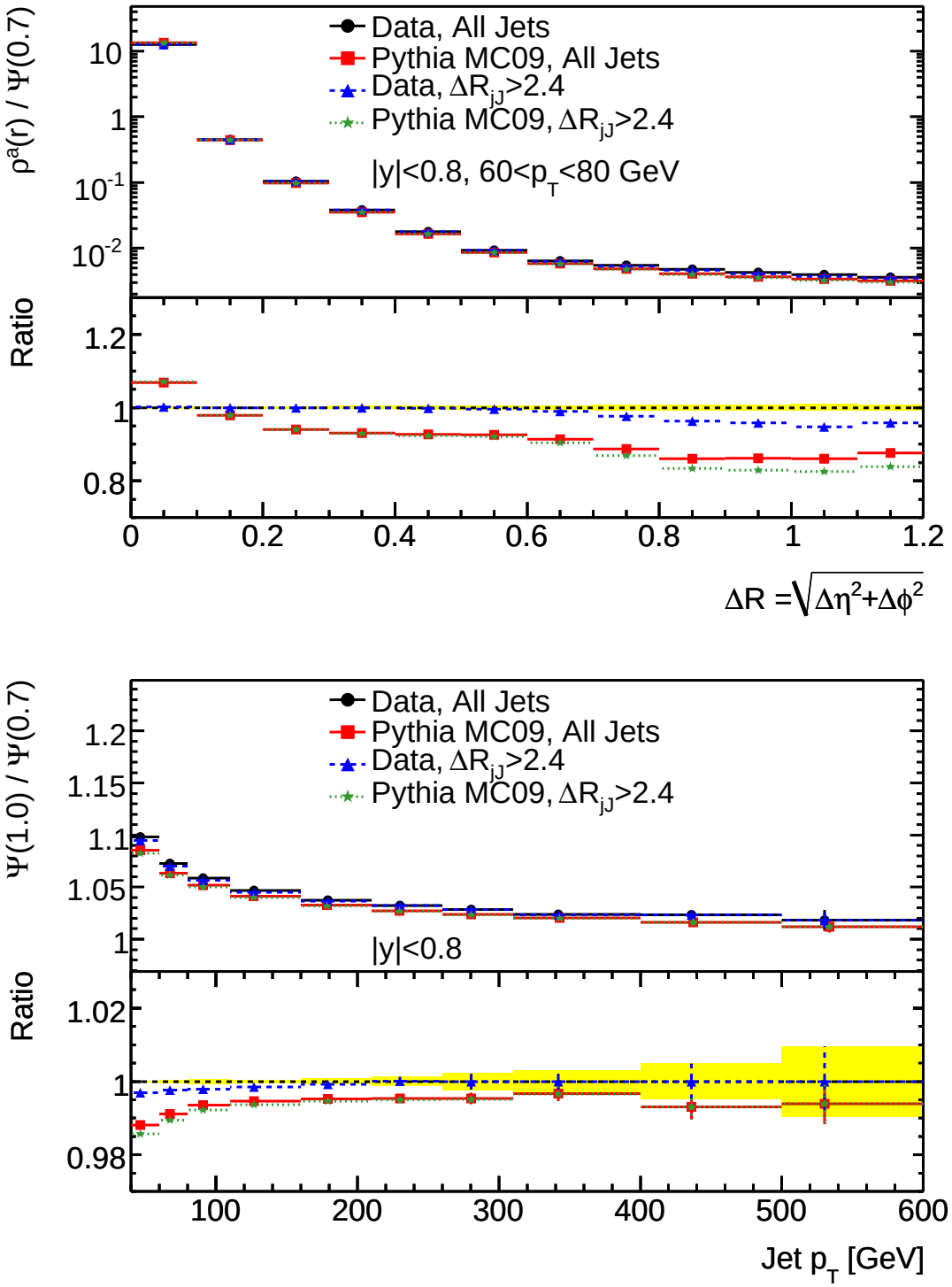


Figure 5.25: Comparison of jet shapes in jets which were far from (ΔR>2.4) any jet of higher p_T and all jets. The ratio is defined with respect to the inclusive sample.

5.6.4 Problematic Detector Regions

During the data-taking period, there were two major problems with the calorimetry that resulted in a loss of information (see Section 4.2). Although these two problems only impacted ∼3% of the calorimetry, the affected areas were distributed throughout the detector and thus included most jets in some way. The problems are particularly important to understand, since the number of dead channels is time-dependent and, therefore, not completely modeled by the Monte Carlo. A correction for the tile calorimeter problem was included in the standard reconstruction, and the correction for the dead OTXs was not. The latter had not been sufficiently validated for the standard reconstruction by the time of this analysis.

In the tile calorimeter, there were several power supply failures. In these areas, and also in the case of isolated dead cells, the cell energy was approximated using the neighboring (good) cells. The fraction of jet energy due to this correction, for both failures in the tile calorimeter and isolated dead cells, is shown in Figure 5.9. The agreement between data and Monte Carlo simulation is good, and less than 0.1% of jets have a correction larger than 20%. The correction is not sufficient to recover lost jets (i.e. jets where most energy was deposited in a problematic region). Therefore, Figure 5.10 shows some shadows with a deficit of good jets.

In the LAr calorimeter, several optical transmitters (OTXs) failed. By default, no correction was applied to the calorimeter reconstruction to account for this effect. Thus, in the distributions of good jets (Figure 5.10), there are several "cold" spots, where OTXs in important layers of the LAr calorimeter were disabled. Because the level one calorimeter trigger read out is completely analog, it provides an alternate source of data for these cells. For this study, ∼10 nb^{-1} of the data were reconstructed with a correction applied. The energies of cells in regions with a dead OTX were replaced using information from the level one calorimeter trigger. The information from the trigger has a granularity of 0.1×0.1 in $\eta - \phi$ ("trigger towers"), and the measurements suffer from considerably more noise than the standard calorimeter read out. Therefore, the correction was only used for towers in which the energy difference between the trigger measurement and the offline measurement was greater than 2 GeV. The energy from the trigger tower was smeared evenly over all cells in the tower that were affected by a dead OTX. The distribution of good jets in η and ϕ after this correction is shown in Figure 5.26. The "cold" regions that were present before

are much less prominent, although the shadows from the tile calorimeter problems are still present.

The distribution of the fraction of jet energy from the correction for dead optical transmitters is shown in Figure 5.27. About 10% of jets include some correction, and 0.5% of jets are corrected by more than 20%. Because the OTX correction uses an alternate read-out chain, it is more reliable than a nearest-neighbor correction. But, because the energy is smeared evenly over all cells in a tower, the jet shapes may also be smeared.

In order to evaluate the quality of these corrections, jet shapes were compared between those jets with a relatively large correction and all jets. In the case of the tile calorimeter correction, because the neighboring cells are used, it is possible that a fluctuation up or down in the energy of the jet is completely missed. In the case of the LAr calorimeter correction, fluctuations should be taken into account, albeit with a reduced position and energy resolution.

Considering either the OTX or tile calorimeter correction, jets with a large fractional corrections are significantly narrower than the inclusive sample of jets (cf. Figure A.20 and A.21). The correction, however, does not significantly affect the mean of the distributions. The narrower jets in the two corrections can be explained by selection effects. The OTX failures will cause a narrow jet to be completely lost, and by applying the correction it can be recovered. The energy drop profile of a jet is non-linear, and because of the non-linearities a neighboring cell correction assigns too much energy to cells near the center of a jet. Requiring a large correction selects jets where these effects are particularly important. In both cases, the Monte Carlo simulation models well both the fraction of jets with large corrections and the magnitude of the difference between jets with large corrections and all jets.

Both these detector problems were asymmetric in the detector in η and ϕ. Thus, differences in jet shapes on either side of the detector in y and in ϕ provide an additional test of biases due to detector issues. Significant changes from one side of the detector to the other that are not modeled in the Monte Carlo simulation would indicate biases in the unfolded distributions. This comparison also ensures that binning in the magnitude of the rapidity is safe. The differences between the four sub-samples and the inclusive sample are all consistent with statistical variations (cf. Figures A.22 and A.23 and see the beginning of this chapter for a note on statistical uncertainties in weighted Monte Carlo samples).

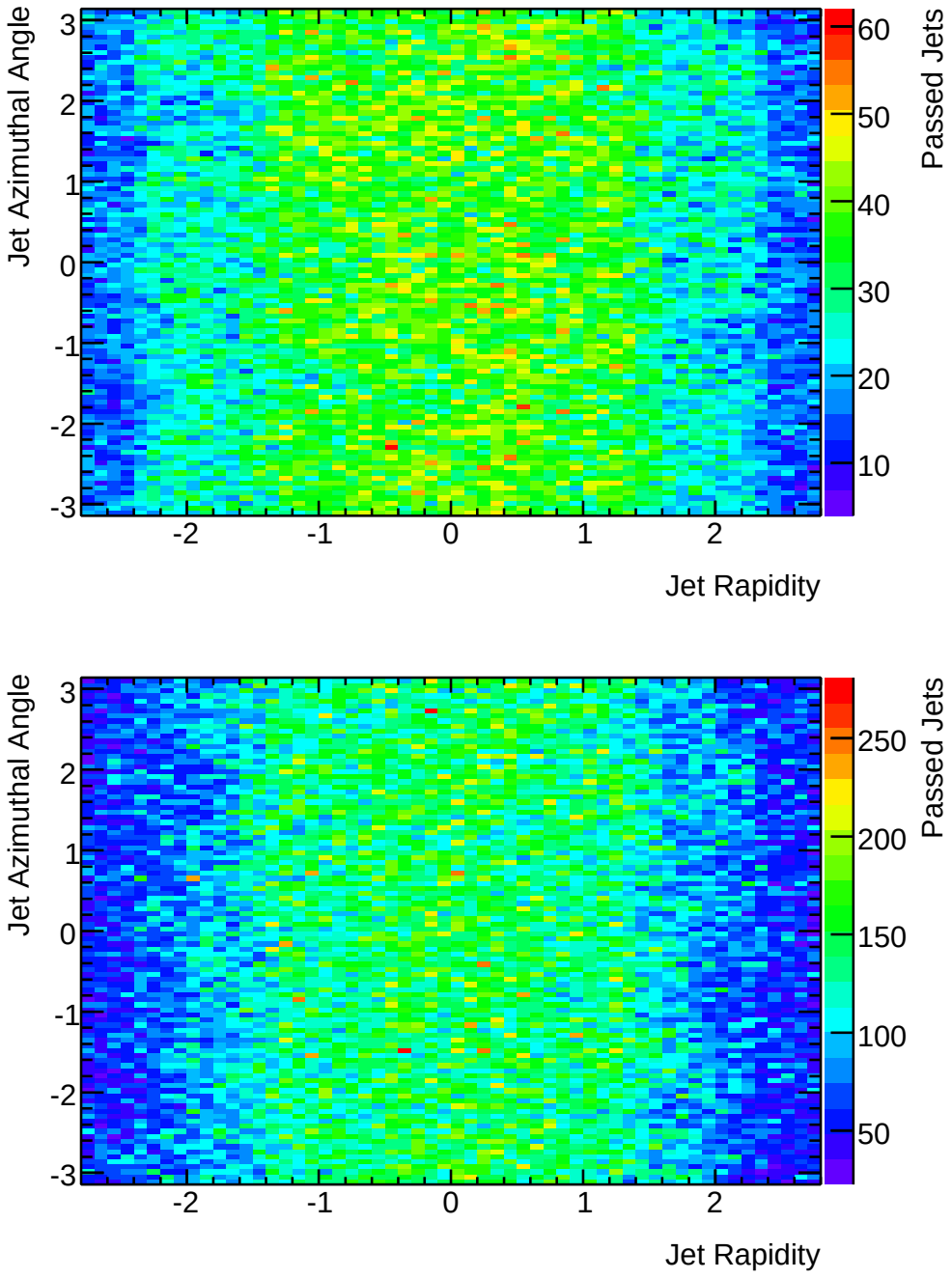


Figure 5.26: Distributions of good jets in data (top) and Monte Carlo simulation (bottom) after correcting for dead optical transmitters in the LAr calorimeter. The "cold" regions that were present without the correction are heavily suppressed.

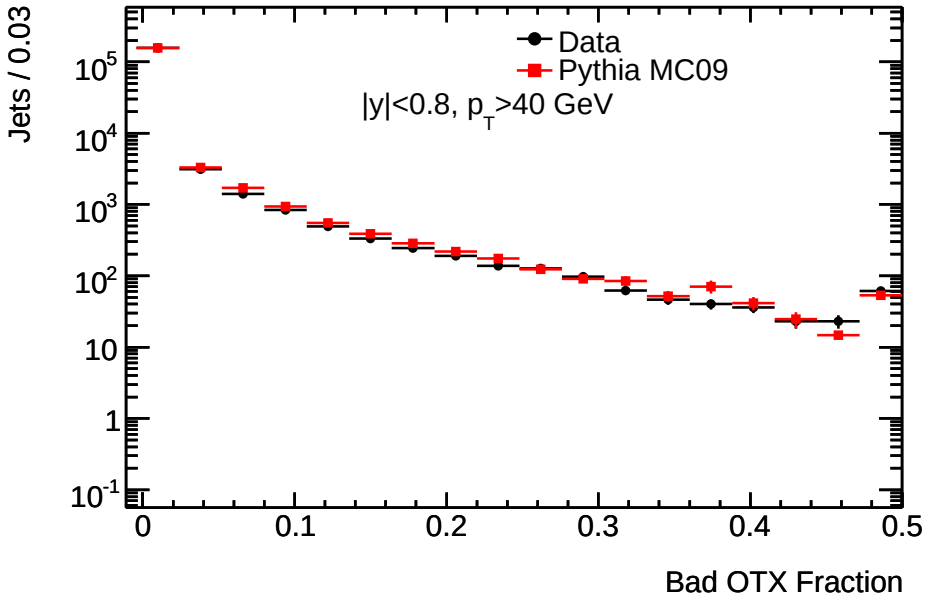


Figure 5.27: Fraction of the jet energy provided by the correction for the dead optical transmitters. The far right bin includes overflow.

5.6.5 Correction for Additional Proton-Proton Collisions

As the luminosity of the LHC increases, events with high-p_T jets and a single vertex will become rare. Particularly in searches for new physics, requiring exactly one reconstructed vertex will result in an unacceptable reduction of statistics. The differences between jets in events with one vertex and two vertices was discussed briefly in Section 5.4.3. It is clear from that discussion that jets receive some contribution on average from the additional proton-proton collisions and that this contribution significantly alters the jet shapes near the edges of the jet. Therefore, it is useful to develop some method for correcting jet energy and jet shape in events with several vertices.

Such a method for correcting the energy of jets has already been developed in ATLAS [188]. The average energy deposition in a single calorimeter tower increases as a function of the number of good reconstructed vertices in the event. This "offset" to the typical energy in a calorimeter tower can be removed from each calorimeter tower in a jet, providing, on average, a subtraction of the energy coming from the additional proton-proton collisions in the event. This offset correction can be applied to jet shapes as well, simply by constructing the jet shape using the offset-corrected calorimeter towers.

Figure 5.28 shows a comparison of jet shapes before and after the offset correction has been applied. In two-vertex events, the difference of 20% at the edge of the jet and 40% outside the jet is reduced to below 5%. In three-vertex events, the difference of 30% at the edge of the jet and 70% outside the jet is reduced to below 10%.

On a jet-by-jet basis, an average offset correction may not behave properly. Some jets receive a larger contribution from additional proton-proton collisions than others. For a statistical analysis of jet shapes, this offset correction appears to significantly reduce the effect of additional proton-proton collisions. An event-by-event discriminant, however, might benefit more from a correction based on the fraction of tracks pointing towards the jet that originate in the primary vertex.

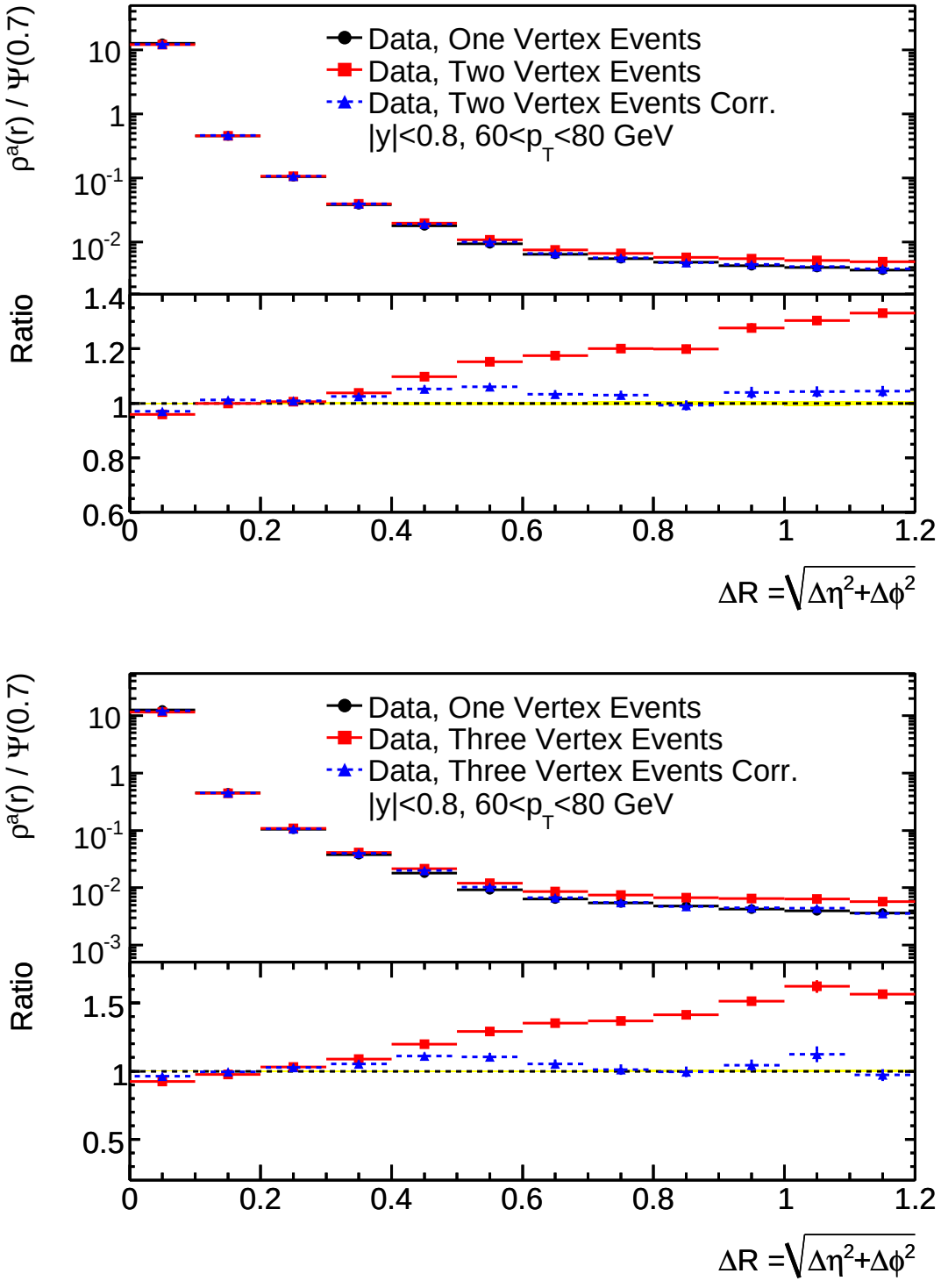


Figure 5.28: Comparison of jet shapes in events with one and two (top) or one and three (bottom) vertices, before and after an offset correction has been applied. The ratio is defined with respect to the single-vertex sample.

Chapter 6 Results

Raw distributions of the $\rho(r)$ variable for several values of r and for several jet p_T bins are shown in Figures 6.1 and 6.2. In these plots, $\rho(r)/\Psi(0.7)$ is ten times (by convention ρ includes a factor of $1/\delta r$) the fraction of jet p_T within an annulus around r (i.e. there is no normalization to the area of the annulus, as there is in ρ^a). There are significant variations jet to jet, due to fluctuations in fragmentation and patron showering as well as hadronic showering within the calorimeter. The same distributions from the Monte Carlo simulation, both at the hadron level and the calorimeter level, are overlaid. Even at this level it is apparent that the data are somewhat wider than the Monte Carlo simulation. The distributions of $\rho(0.05)$ and $\rho(0.15)$ are somewhat higher in the data.

The variation from jet-to-jet typically implies a large-scale difference in jet structure. That is, jets with a small $\rho(r)/\Psi(0.7)$ typically display a bi-nodal structure. Frequently, such jets would be split into two by jet algorithms with a smaller jet-finding radius. On the other hand, jets with a large $\rho(r)/\Psi(0.7)$ are typically collimated jets, and often have one or two hard leading hadrons that contribute a large fraction of the jet energy. Far from the center of the jet, the jet shape is dominated by the underlying event and calorimeter noise. In the calorimeter jet shapes, the additional smearing due to the hadrons' shower in the calorimeter reduces the number of jets with small $\rho(0.05)/\Psi(0.7)$.

For all results in this section, unfolded distributions with systematic uncertainties will be compared to several different Monte Carlo programs and several tunes of Pythia. A summary of the programs and tunes is given in Table 6.1. Pythia with the MC09 tune was used to derive the unfolding corrections. Alpgen is the only generator that uses a $2 \to n$ matrix element. All tunes of Pythia use the same hadronization and underlying event models, but several include re-tunes of the hadronization and underlying event model parameters. Herwig and Alpgen use the same hadronization model and the underlying event of Jimmy, though with different tunes. Alpgen and the Perugia0, Perugia2010, and DW tunes all used LO PDFs. The other generators and tunes used modified LO PDFs. Herwig++ includes similar underlying assumptions and models to those of Herwig and Jimmy, but there were many modifications and differences in the tunes of the models that

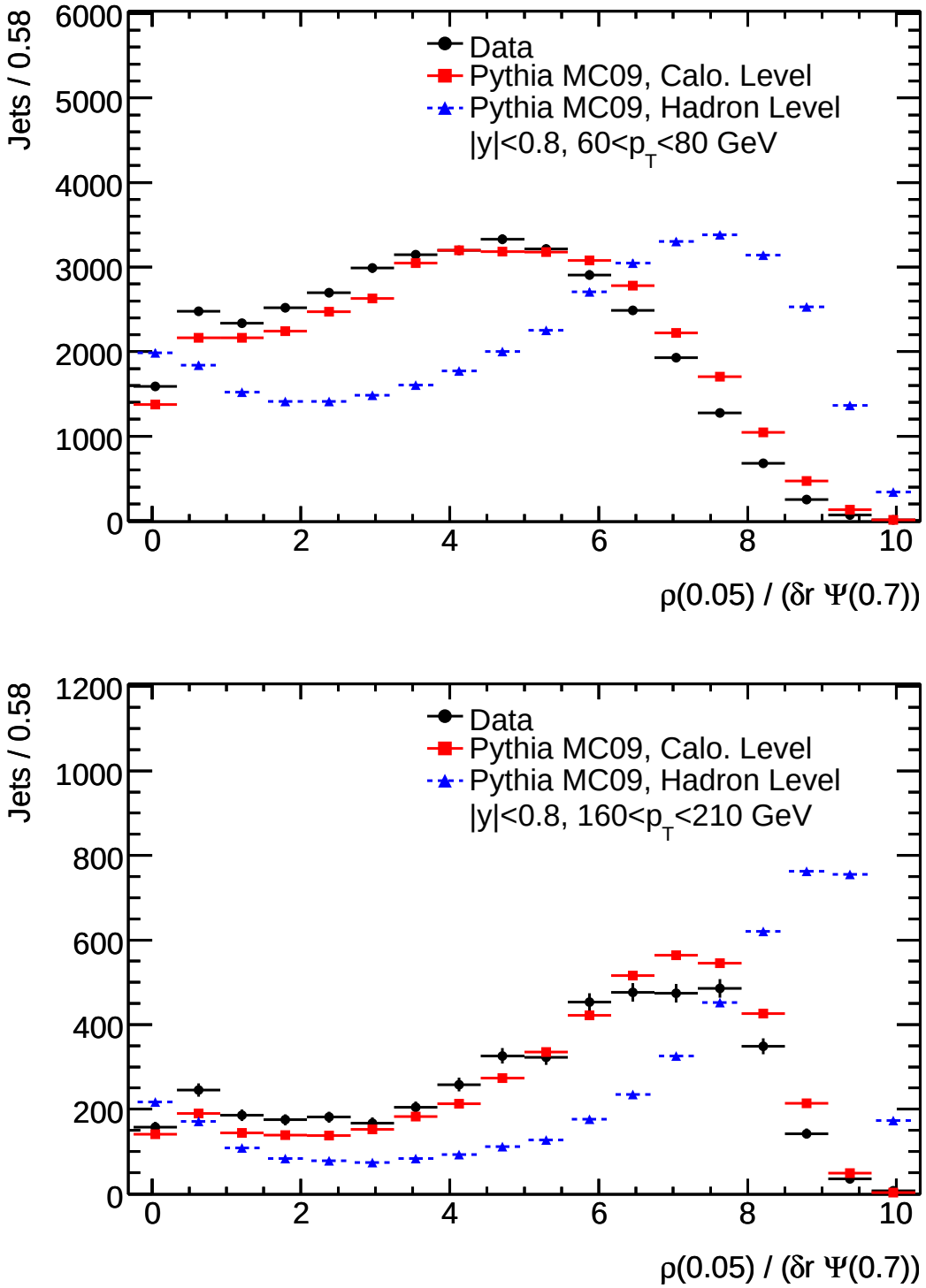


Figure 6.1: Raw distributions of the $\rho\,(0.05)\,/\Psi\,(0.7)$ variable in two jet p_T bins. The factor of $1/\delta r$ is by convention in the definition of the variable, so the range is from 0 to 10.

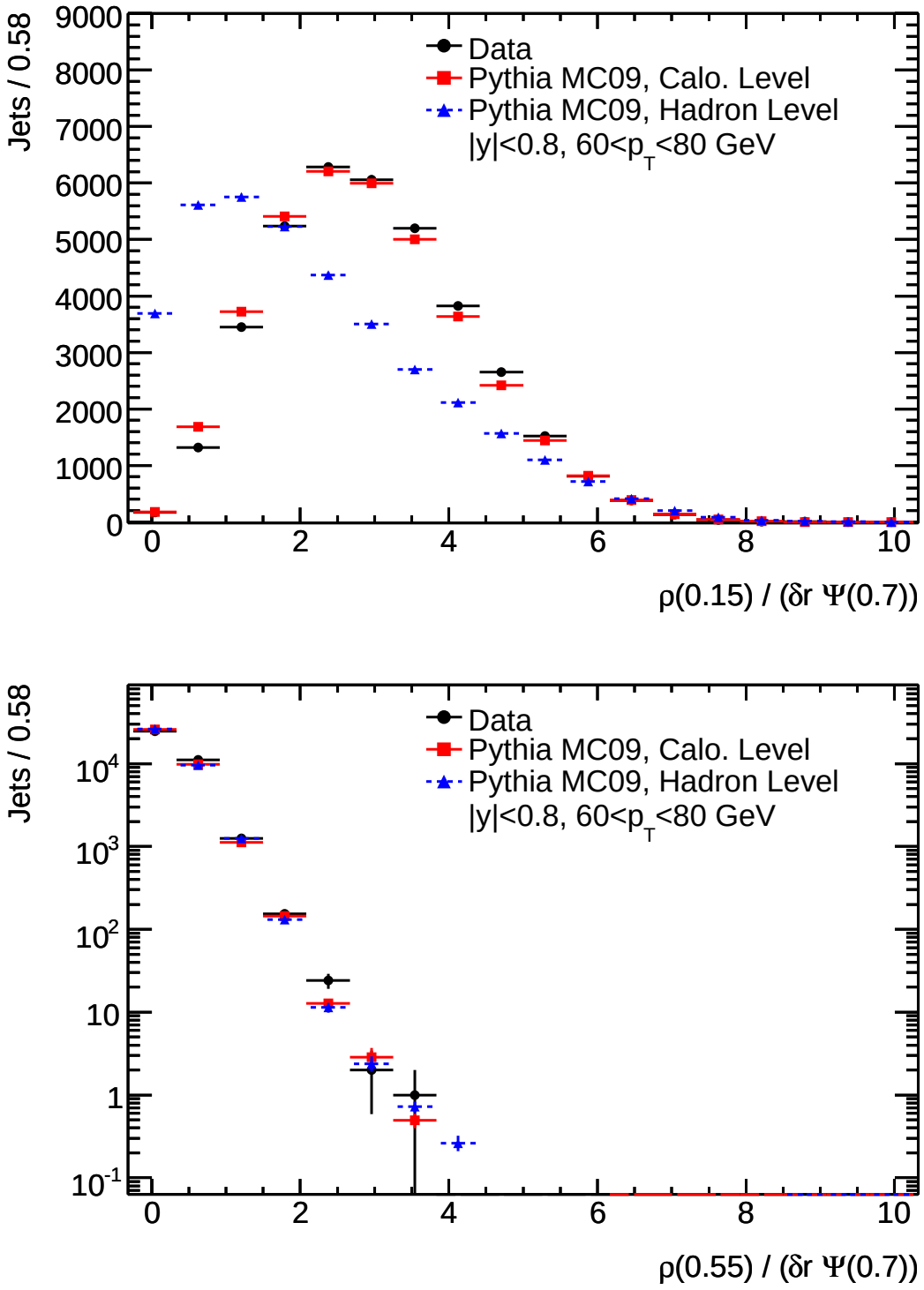


Figure 6.2: Raw distributions of the $\rho(r)/\Psi(0.7)$ variable for $r = 0.15$ and $r = 0.55$. The factor of $1/\delta r$ is by convention in the definition of the variable, so the range is from 0 to 10.

make the results dissimilar.

Program / Tune	2→	Hadr. Model	Underlying Event	PDFs	Reference
Pythia / MC09	2	String	1.0	Mod. LO	[63]
Pythia / Perugia0	2	String	1.1	LO	[64]
Pythia / Professor	2	String	1.0	Mod. LO	[62]
Pythia / AMBT1	2	String	1.2	Mod. LO	[189]
Pythia / Perugia2010	2	String	1.3	LO	[190]
Pythia / DW	2	String	1.4	LO	[191]
Herwig / Jimmy	2	Cluster	2.0	Mod. LO	[56, 57, 163]
Alpgen	n	Cluster	2.0	LO	[42]
Herwig++	2	Cluster	2.5	Mod. LO	[58]

Table 6.1: Key parameters for the Monte Carlo programs and tunes used for comparison in this chapter. The underlying event model numbers indicate differences in underlying assumptions (first number) and differences in tuning or implementation (second number).

After unfolding has been applied, the differential jet shapes, $\rho^a(r)/\Psi(0.7)$ are compared to the Monte Carlo predictions in each rapidity bin and in the four lowest p_T bins (i.e. where there are sufficient statistics for an interesting comparison). Comparisons to different programs are shown in Figures 6.3 and 6.4 (see also Figures A.24, A.25, A.26, A.27, A.28, A.29, A.30, A.31, A.32, and A.33). Comparisons to different tunes are shown in Figures 6.5 and 6.6 (see also Figures A.34, A.35, A.36, A.37, A.38, A.39, A.40, A.41, A.42, and A.43). The shaded bands represent the sum in quadrature of the statistical uncertainties and systematic uncertainties on the data. Only in the highest p_T bin outside of the central region are the statistical uncertainties relevant. At large r, ρ^a has an asymptotic behavior, where the underlying event dominates the shape.

In all cases, the systematic uncertainties are large. However, some differences between the Monte Carlo and the data are visible. In the core of the jet, Pythia with the Perugia2010 tune seems to have the best agreement, though several predictions are within the systematic uncertainties. As Perugia2010 is the only tune that included the Tevatron jet shape measurements, the agreement is not a surprise. Far from the core of the jet, particularly outside the jet "cone," Herwig++ includes too much activity, likely indicating too large an underlying event contribution. As is suggested by other ATLAS analyses, in particular the opening angle between the two leading tracks in a jet [171], the fragmentation in the data appear to produce wider jet cores than most Monte Carlo tunes.

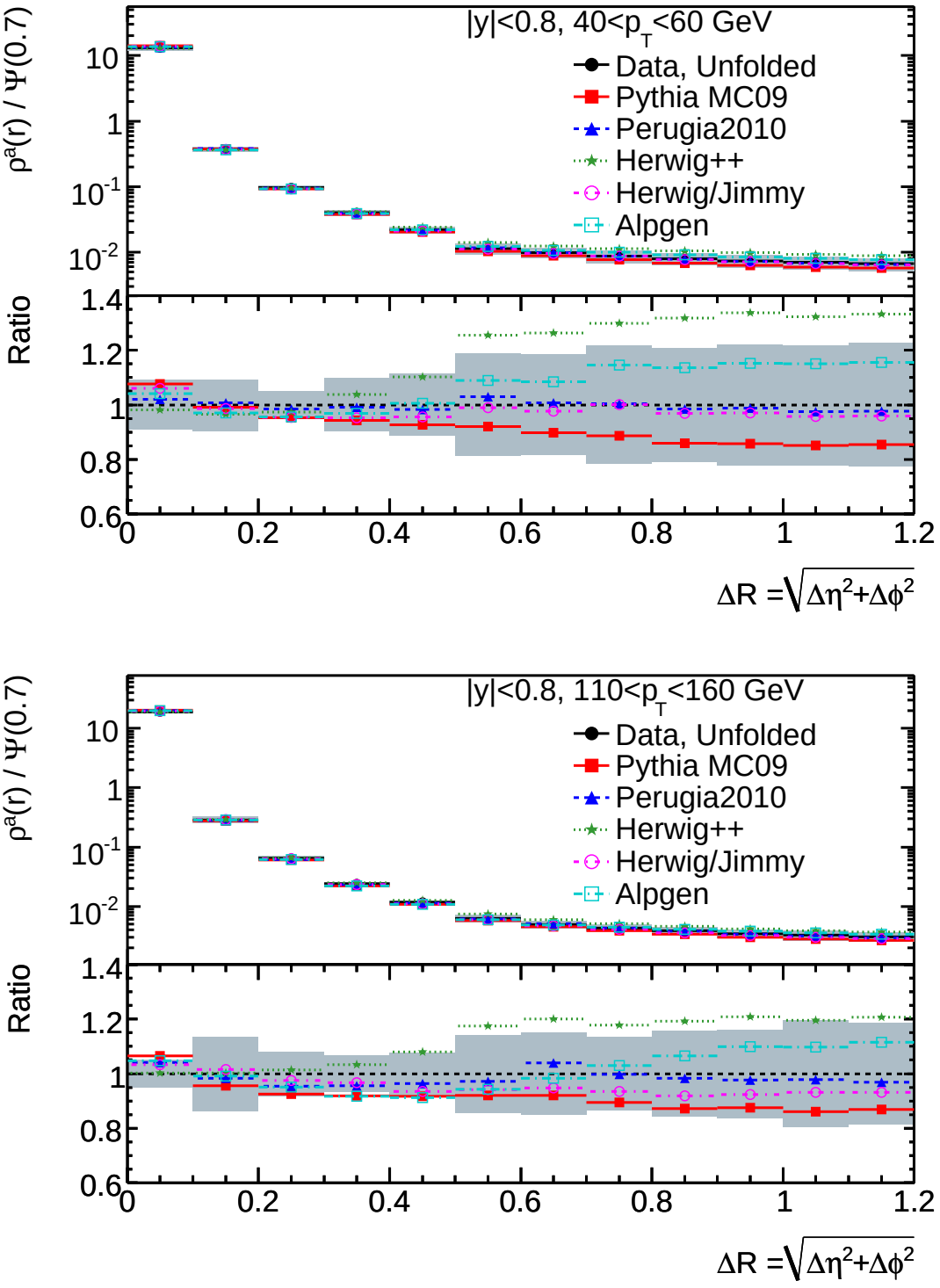


Figure 6.3: Differential jet shapes after unfolding, for jets with $|y| < 0.8$, in two p_T bins.

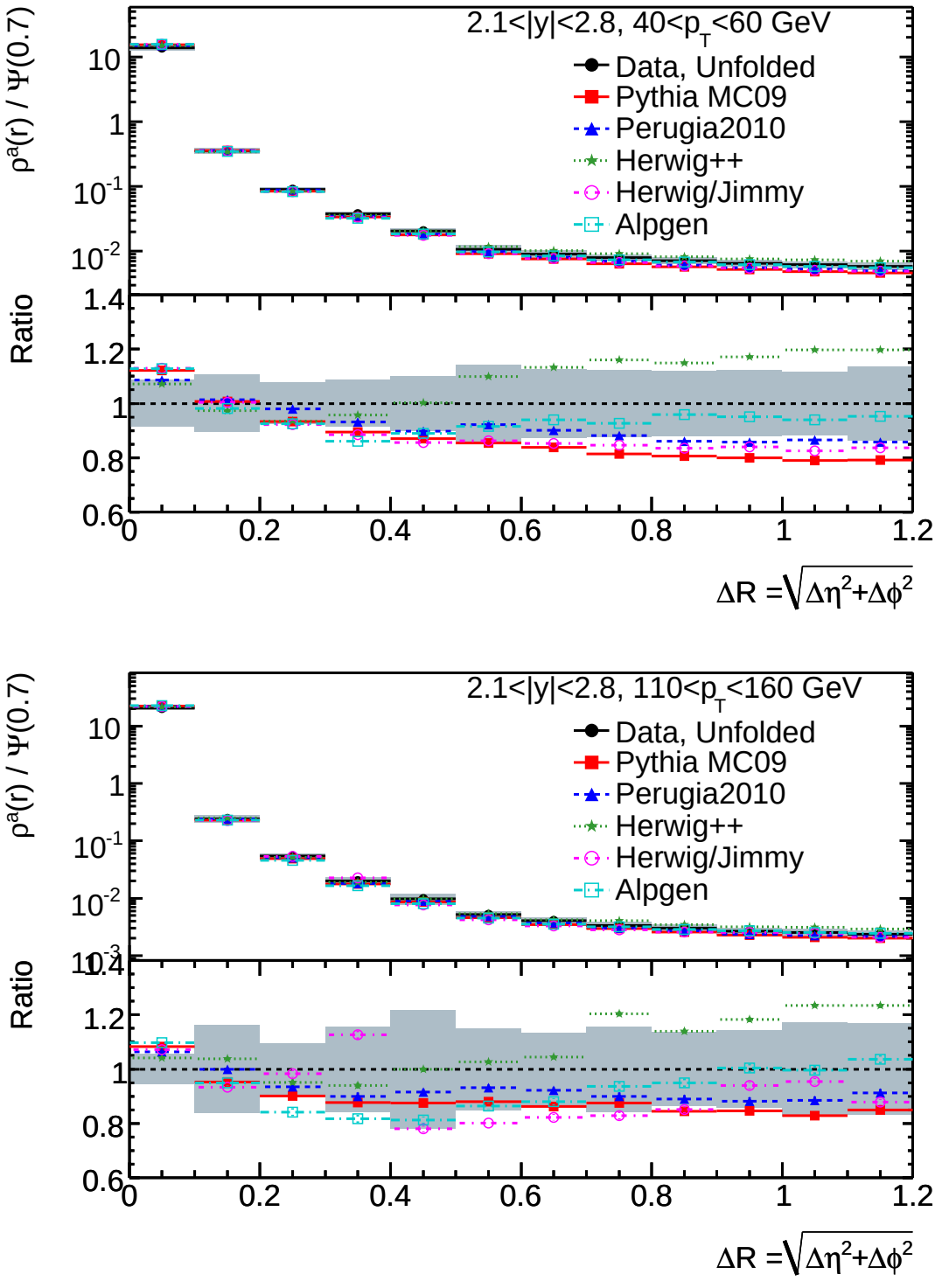


Figure 6.4: Differential jet shapes after unfolding, for jets with $2.1 < |y| < 2.8$, in two p_T bins.

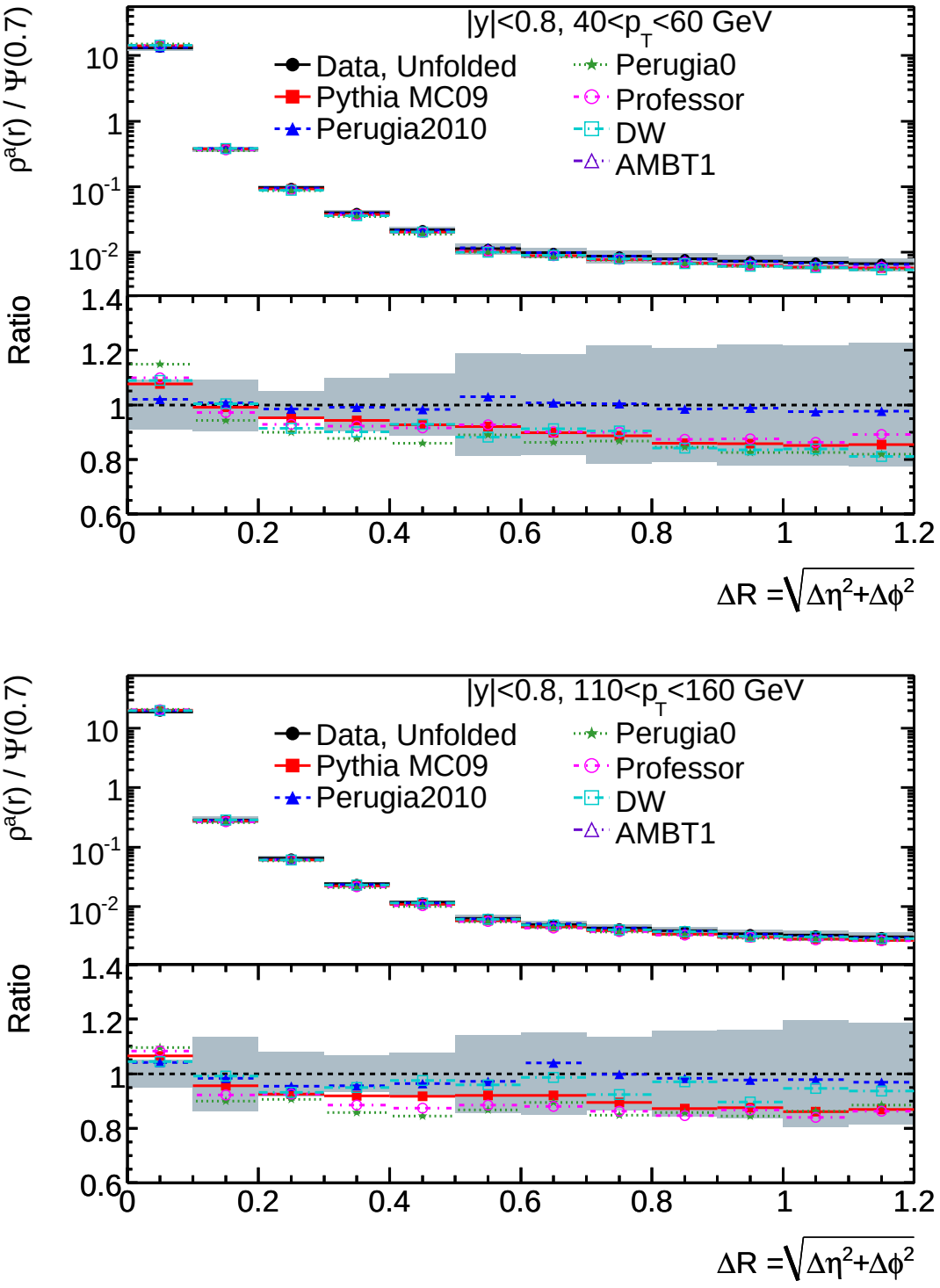


Figure 6.5: Differential jet shapes after unfolding, for jets with $|y| < 0.8$, in two p_T bins.

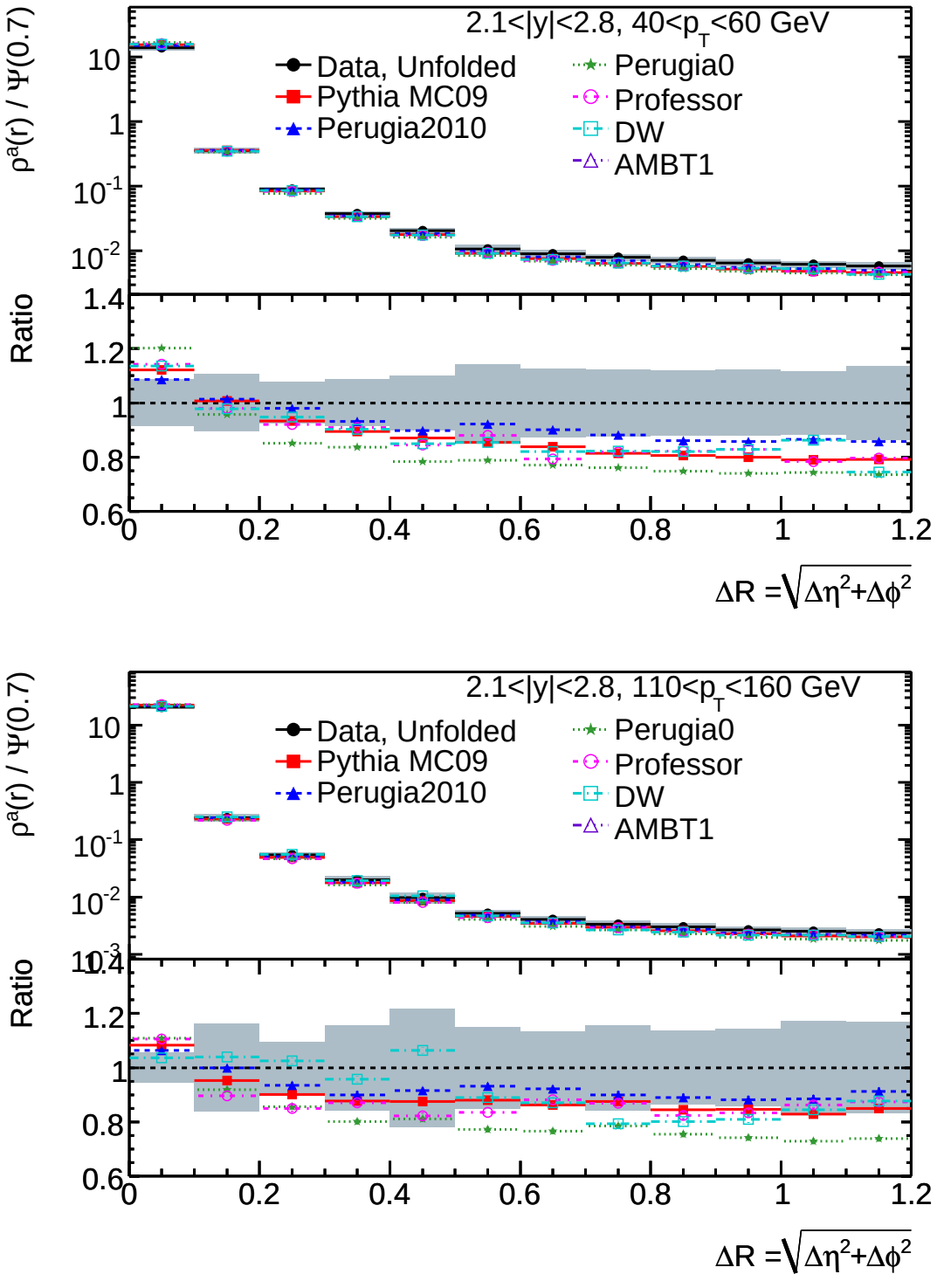


Figure 6.6: Differential jet shapes after unfolding, for jets with $2.1 < |y| < 2.8$, in two p_T bins.

At large rapidity ($2.1 < y < 2.8$), most of the PYTHIA tunes produce too little activity far from the jet, and HERWIG++ still produces too much. ALPGEN seems to provide the best description of the data here. It includes the underlying event model and hadronization model from JIMMY and HERWIG, respectively, but used a different tune and different PDFs for the purpose of this comparison. Thus, the importance of the inclusion of the $2 \rightarrow n$ matrix elements in ALPGEN and the different duplicate removal scheme (MLM rather than CKKW, see Chapter 2) is still ambiguous.

After unfolding has been applied, the p_T-dependent jet shapes, $\Psi(0.3)/\Psi(0.7)$ are compared to Monte Carlo predictions in each rapidity bin. The comparisons to different Monte Carlo programs are shown in Figure 6.7 (see also Figure A.44). The comparisons to different tunes of PYTHIA are shown in Figure 6.8 (see also Figure A.45). The PYTHIA DW tune could not be used for this comparison, as the samples of high p_T jets were not available. Again, the shaded bands represent the sum in quadrature of the statistical and systematic uncertainties on the data. In the highest three p_T bins outside of the central region and the highest p_T bin in the central region, the statistical uncertainties are dominant. There are fewer than ten jets in the highest p_T bin in the rapidity regions $0.8 < |y| < 1.2$ and $1.2 < |y| < 2.1$. Because of the lack of statistics, jet shapes are not shown above 400 GeV in the most forward rapidity region ($2.1 < |y| < 2.8$).

The variation between the various Monte Carlo programs and tunes is only a few percent. Again, the best agreement is with the PYTHIA Perugia2010 tune. Because it describes the core of the jet well, it describes this variable particularly well. In the high-rapidity bins, all tunes and programs except HERWIG++ produce jets that are narrower than the data, though in most cases consistent given the systematic uncertainties. The jet shapes vary by only a few percent over the four rapidity bins considered.

A comparison of the results derived in this thesis to those of a parallel analysis using ATLAS data [192] and the published analysis from CDF [96] is shown in Figure 6.9. The error bars include statistical and systematic uncertainties, added in quadrature. Systematic uncertainties dominate. To ensure a fair comparison, the historical definition of $\rho(r)$ is used for this comparison. The disagreements between this analysis and the parallel ATLAS analysis are explained by different choices in shape construction. For example, this analysis included in the shape definition all constituents within some radius of the jet axis, whereas the other analysis included only those constituents assigned to the jet by the jet algorithm.

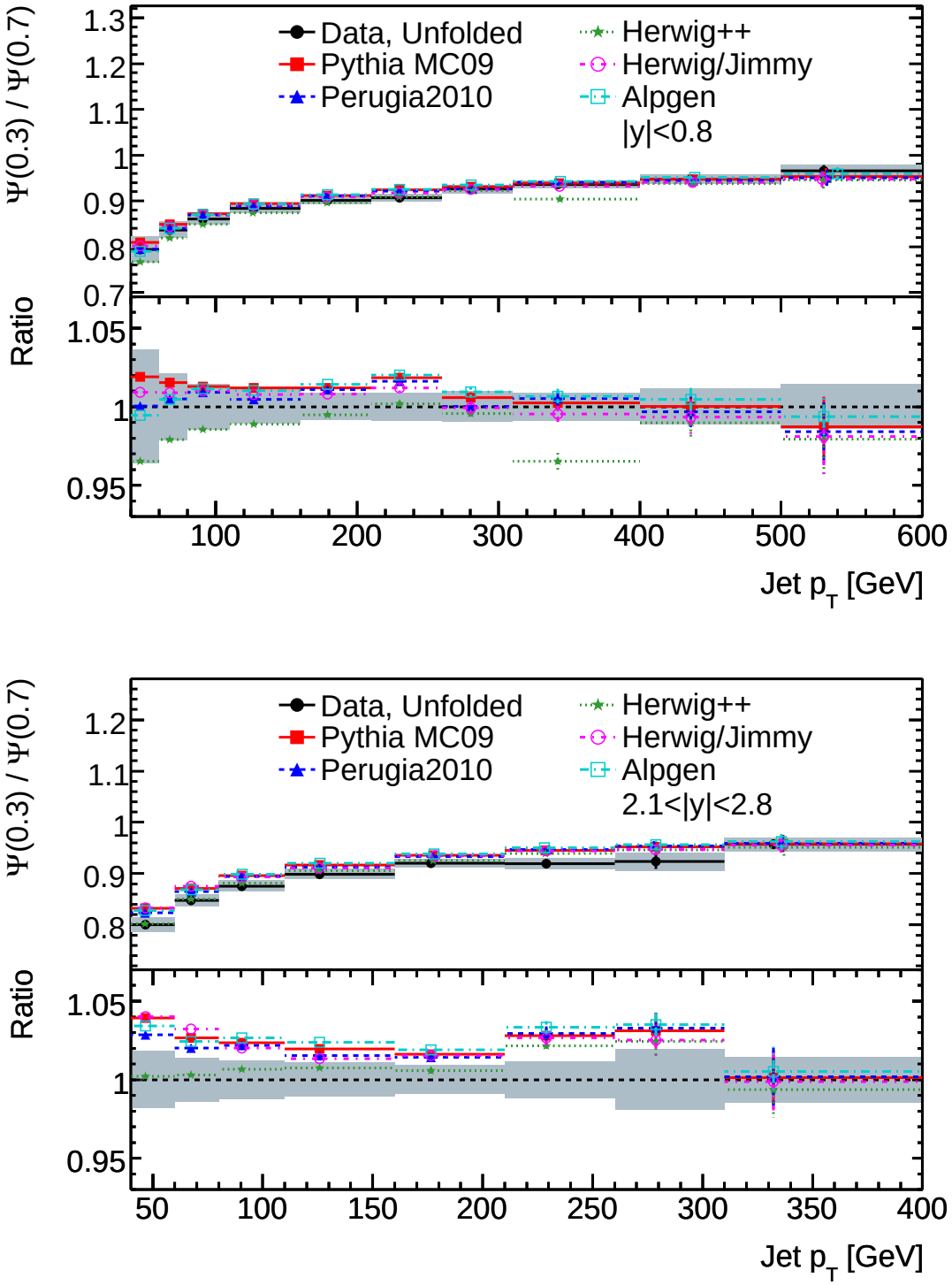


Figure 6.7: Integrated jet shapes after unfolding as a function of jet p_T for jets with $|y| < 0.8$ (top) and $2.1 < |y| < 2.8$ (bottom).

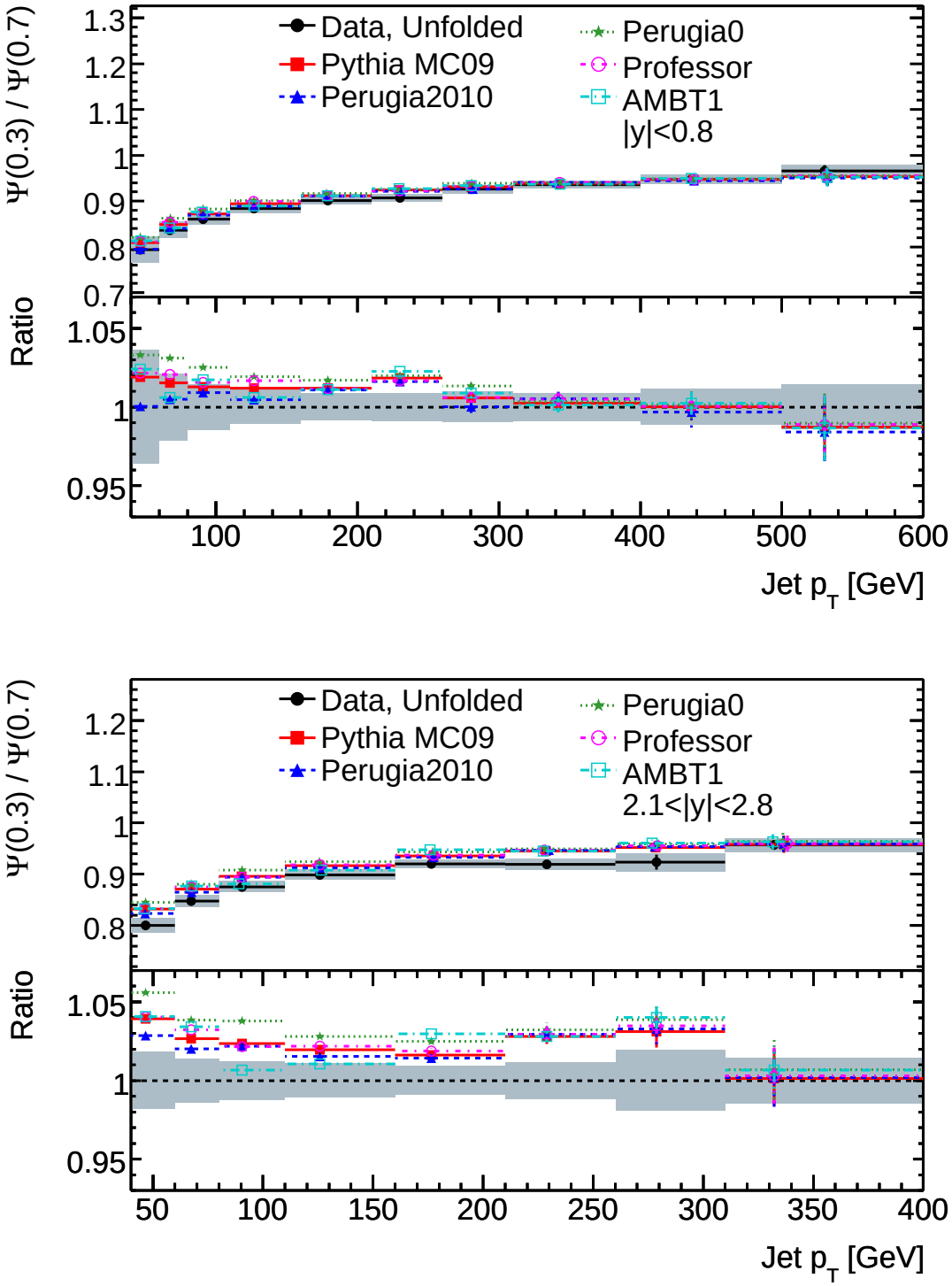


Figure 6.8: Integrated jet shapes after unfolding as a function of jet p_T for jets with $|y| < 0.8$ (top) and $2.1 < |y| < 2.8$ (bottom).

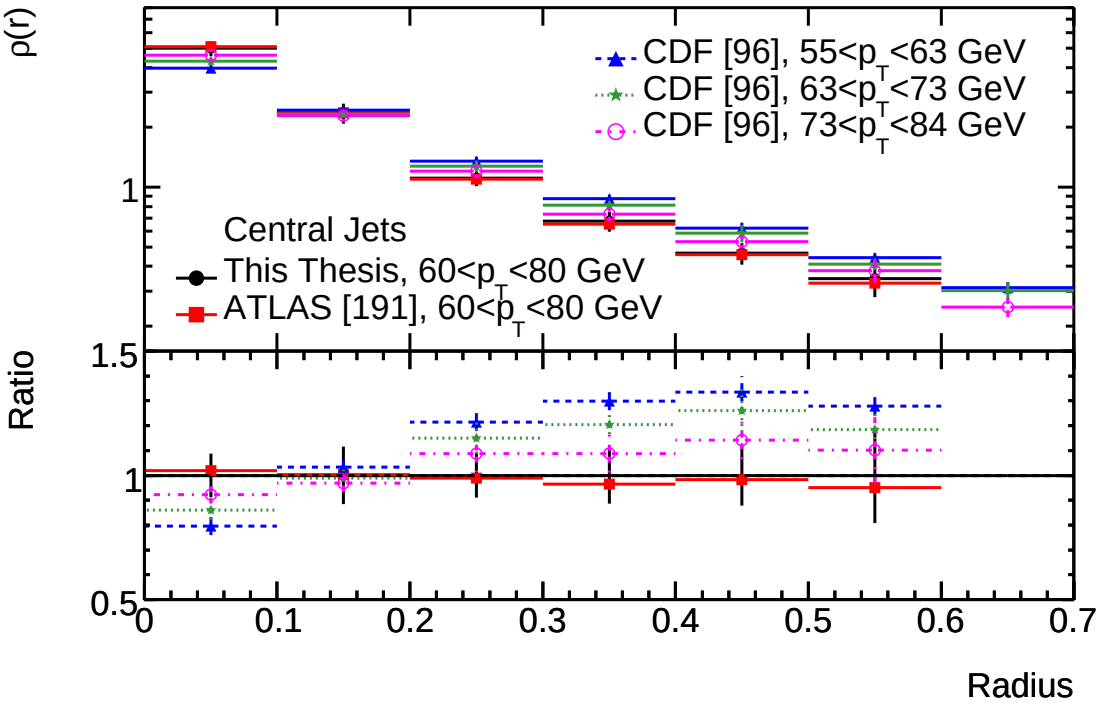


Figure 6.9: A comparison of unfolded jet shapes from this analysis, a parallel analysis in ATLAS, and the published analysis from CDF. Error bars include statistical and systematic uncertainties added in quadrature. No error bars are shown on the alternate ATLAS analysis.

A difference appears when two jets are close to one another. The difference was explored in Section 5.6.3 and is several percent far from the core of the jet. The other ATLAS analysis used topological clusters as inputs to the jet shapes, whereas this analysis used calorimeter towers with topological noise suppression. Because of their geometric area, calorimeter towers are more stable against pile-up effects. Topological clusters gather more energy at the center of the jet (see Section 5.6.2), and therefore present smaller unfolding corrections (though not necessarily smaller unfolding systematic uncertainties). Because $\rho(r)$ is normalized to 10, any differences are always reflected in at least two bins of the variable. Any remaining differences between the results of the two analyses are consistent with the variance in unfolding observed in Chapter 5.

The jets in ATLAS, from collisions at 7 TeV center-of-mass energy, are significantly narrower than those of the same p_T in CDF, from collisions at 1.96 TeV center-of-mass energy. This change with center-of-mass energy is fairly well described by the Monte Carlo

programs considered and comes as a result of several competing effects. At the same p_T, there are fractionally more gluon jets in ATLAS, which would tend to broaden the jet shape distributions. The underlying event is also more significant at the LHC, which would further broaden jet shape distributions in ATLAS. However, the midpoint cone jet algorithm employed by the CDF analysis [74] includes a splitting and merging stage based on the fraction of jet momentum shared by two nearby jets. Because of the rapidly falling p_T spectrum, in a single bin of p_T it is more likely for a jet to have been affected by merging than by splitting. The merging algorithm will result in considerably wider jets at the Tevatron, as are observed. The differences in jet shapes among various jet algorithms is discussed further in Section 2.2. Once more data are available and the Monte Carlo programs have been re-tuned to the LHC data, ATLAS may attempt to calibrate several different jet algorithms. At that point the differences could be studied further.

Some comparisons were made with the available CMS data [193]. However, unfolded results with systematic uncertainties were unavailable at the time of this writing. The measured shapes appear to be considerably more narrow than those of ATLAS, but as no full analysis is available it is not possible at this stage to say what might be causing this effect. The CMS analysis uses a slightly larger jet for their jet shapes studies ($D = 0.7$, rather than 0.6) and used a different Monte Carlo tune for the bin-by-bin unfolding. As the results are highly dependent on the Monte Carlo program used for the unfolding, a large part of the difference in final results may originate there.

6.1 Jet Shapes in Quark- and Gluon-Jets

At the Tevatron, it was possible to overlay the shapes of quark-jets and gluon-jets and demonstrate that the data lie between them [96, 97, 98]. In order to form a set of "gluon-jets" and "quark-jets," hadron-level jets in Pythia were matched to the partons in the event. If a quark was the hardest parton near the jet (i.e. with $\Delta R_{\text{jet,parton}} < 0.7$), the jet was labeled a "quark-jet." If a gluon was the hardest parton near the jet and, in the case of g→qq, both the quarks from the gluon were also near the jet, the jet was labeled a "gluon-jet." This algorithm somewhat favors tagging a jet as a "quark-jet" by neglecting the possibility of hard gluon radiation from quarks.

By this analysis, the jets in the ATLAS data are wider than either quark- or gluon-jets

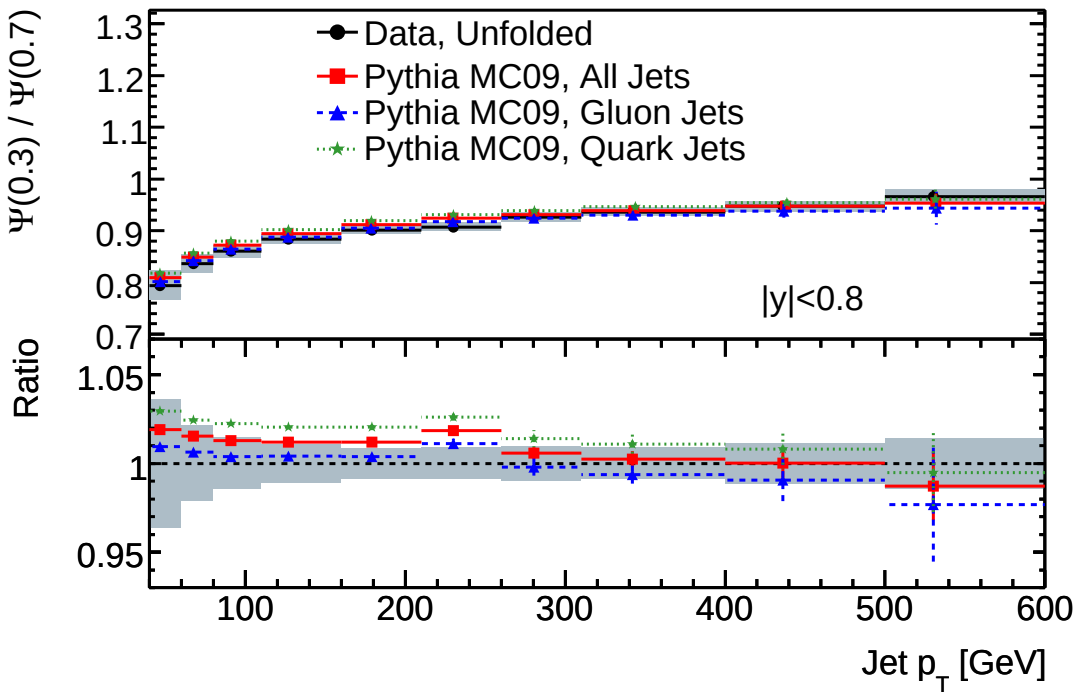


Figure 6.10: A comparison of quark-jets and gluon-jets in Pythia Monte Carlo to jets in data using the jet shape variable $\Psi(0.3)/\Psi(0.7)$.

in Pythia (though consistent with both within systematic uncertainties). A comparison of the jet shape variable $\Psi(0.3)/\Psi(0.7)$ between the data, Pythia with the ATLAS MC09 tune, and quark-jets and gluon-jets from that Monte Carlo sample is shown in Figure 6.10. The variation between quark-jets and gluon-jets is only a few percent, comparable to the total systematic uncertainty on the shape, and is smaller at high p_T.

Other jet shape variables can be constructed that show a more significant variation between quark- and gluon-jets. For example, $(\Psi(0.7) - \Psi(0.1))/\Psi(0.7)$ shows a ~20% variation at low p_T and a larger variation at high p_T. This variable is another measure of jet width. Unfolding corrections and systematic uncertainties were derived for this jet shape in the same way as for $\Psi(0.3)/\Psi(0.7)$. The result of the comparison between quark- and gluon-jets and the data is shown in Figure 6.11. By this measure, the data are inconsistent with this sample of Pythia quark-jets and show good agreement with the Pythia gluon-jets. The differences are more pronounced at high rapidity, where the data agree best with the mixture of quark- and gluon-jets provided by Pythia.

Another similar measure is the jet shape variable $(\Psi(0.1) - \Psi(0.2))/\Psi(0.7)$. This vari-

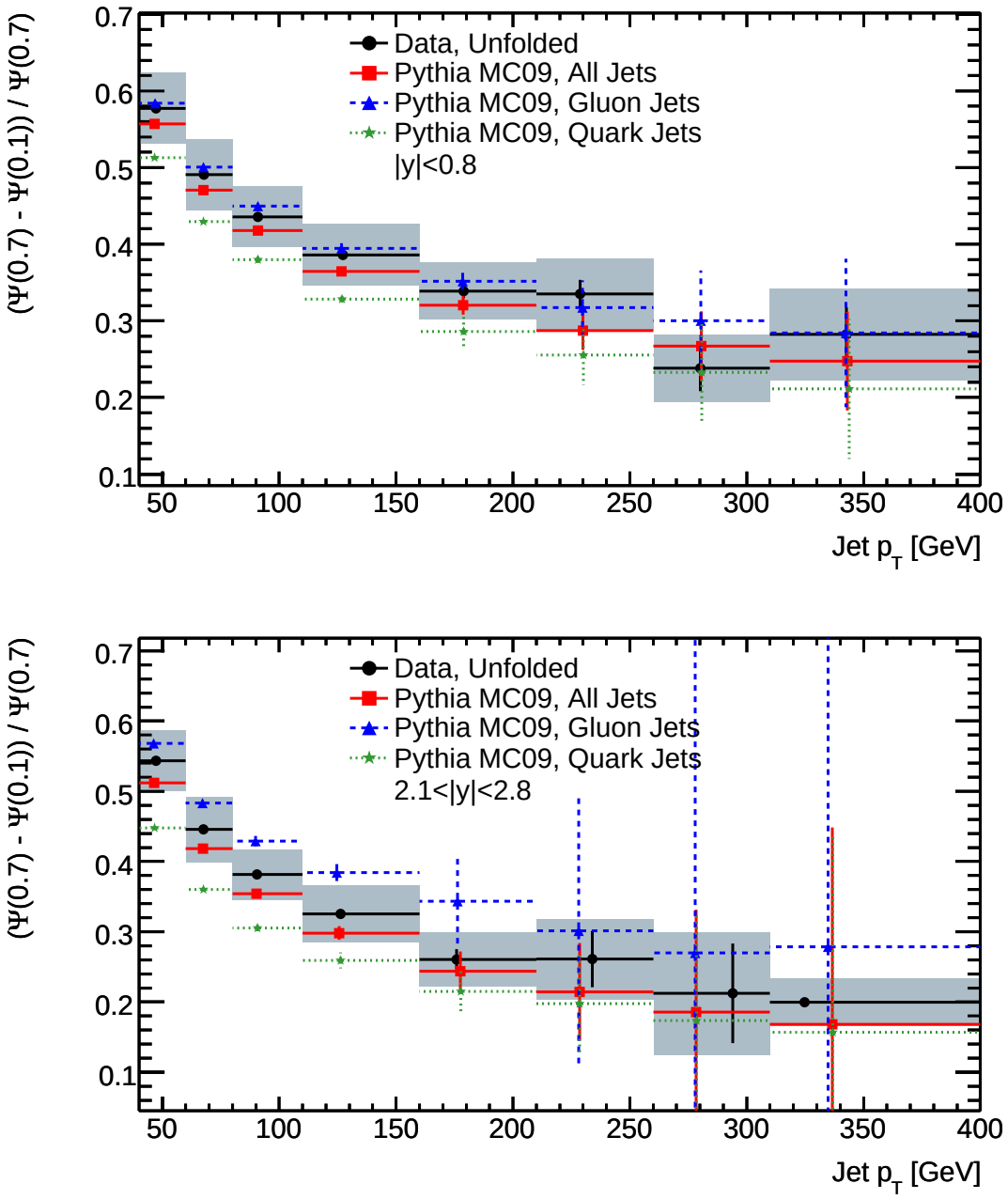


Figure 6.11: A comparison of quark-jets and gluon-jets in PYTHIA Monte Carlo to jets in data using the jet shape variable $(\Psi(0.7) - \Psi(0.1))/\Psi(0.7)$. Top, at low rapidity. Bottom, at high rapidity.

able shows large variation between quark- and gluon-jets at low p_T (up to a factor of two). Unfolding corrections and systematic uncertainties were derived for this jet shape in the same way as the previous one. The result of the comparison between quark- and gluon-jets and the data are shown in Figure 6.12. By this measure, the data are inconsistent with this sample of pure Pythia quark-jets, particularly at high rapidity. The jets in the data show reasonable agreement with the Pythia gluon-jets.

These differences should not be taken to mean that the jets in the data are, in fact, a pure sample of gluon-jets. Rather, it demonstrates the potential for these variables to evaluate the average quark-content of a sample once the inclusive sample shapes agree well. Because of their good discrimination power, they may be valuable in the future in attempts to build up a likelihood variable for distinguishing quark-jets from gluon-jets. With a reduction of systematic uncertainties, it may also be possible to evaluate the differences between various PDFs.

6.2 Jet Shapes in Sub-samples

Because the simplest picture of proton-proton scattering involves a 2→2 process, it is interesting to compare jet shapes for the leading jet to those of the sub-leading jets. Generators like Pythia, Herwig, and Herwig++ include approximations that fail when there are three jets in the event of comparable p_T (see Chapter 2). Alpgen, on the other hand, models the 2→n processes directly at the matrix element level, without the approximation that the third leading jet is from radiation. It may, therefore, give a better description of the third leading jet at high p_T.

In order for these comparisons to be fair, the kinematics of the sub-leading jets must be similar in the data and the Monte Carlo. Otherwise, the kinematic differences would muddle any conclusions about real differences in the jet shapes. The transverse momentum distributions for the third- and fourth-leading jets agree to within the statistical uncertainties (cf. Figure A.46), so the difference in kinematics should have at most a small effect on the unfolded shape distributions.

Figure 6.13 compares the shapes for the leading and second leading jet after unfolding to those in various Monte Carlo generators. Figure 6.14 compares the shapes for the third and fourth leading jet after unfolding to those in various Monte Carlo generators. The

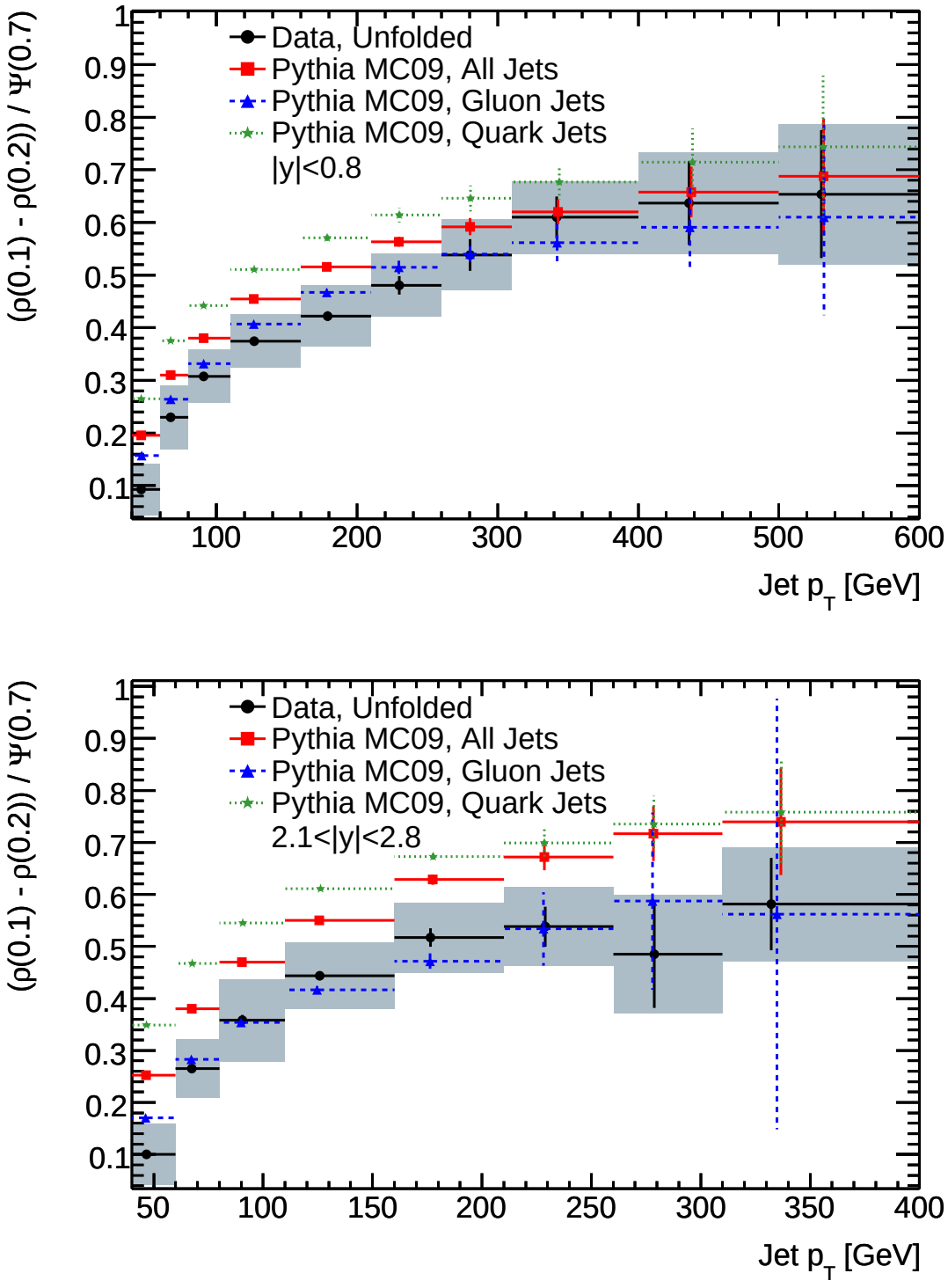


Figure 6.12: A comparison of quark-jets and gluon-jets in PYTHIA Monte Carlo to jets in data using the jet shape variable $(\Psi(0.1) - \Psi(0.2))/\Psi(0.7)$. Top, at low rapidity. Bottom, at high rapidity.

generators describe the shapes of the sub-leading jets well over the available p_T range. Once they have been re-tuned and the systematic uncertainties have been reduced, it may be more possible to draw some conclusion about the differences between the shapes of the leading and sub-leading jets. Jet shapes were also tested in events where exactly two jets were reconstructed with p_T>40 GeV and in y<2.8. The agreement between the unfolded data and the Monte Carlo generators in these events showed no significant improvement or degradation.

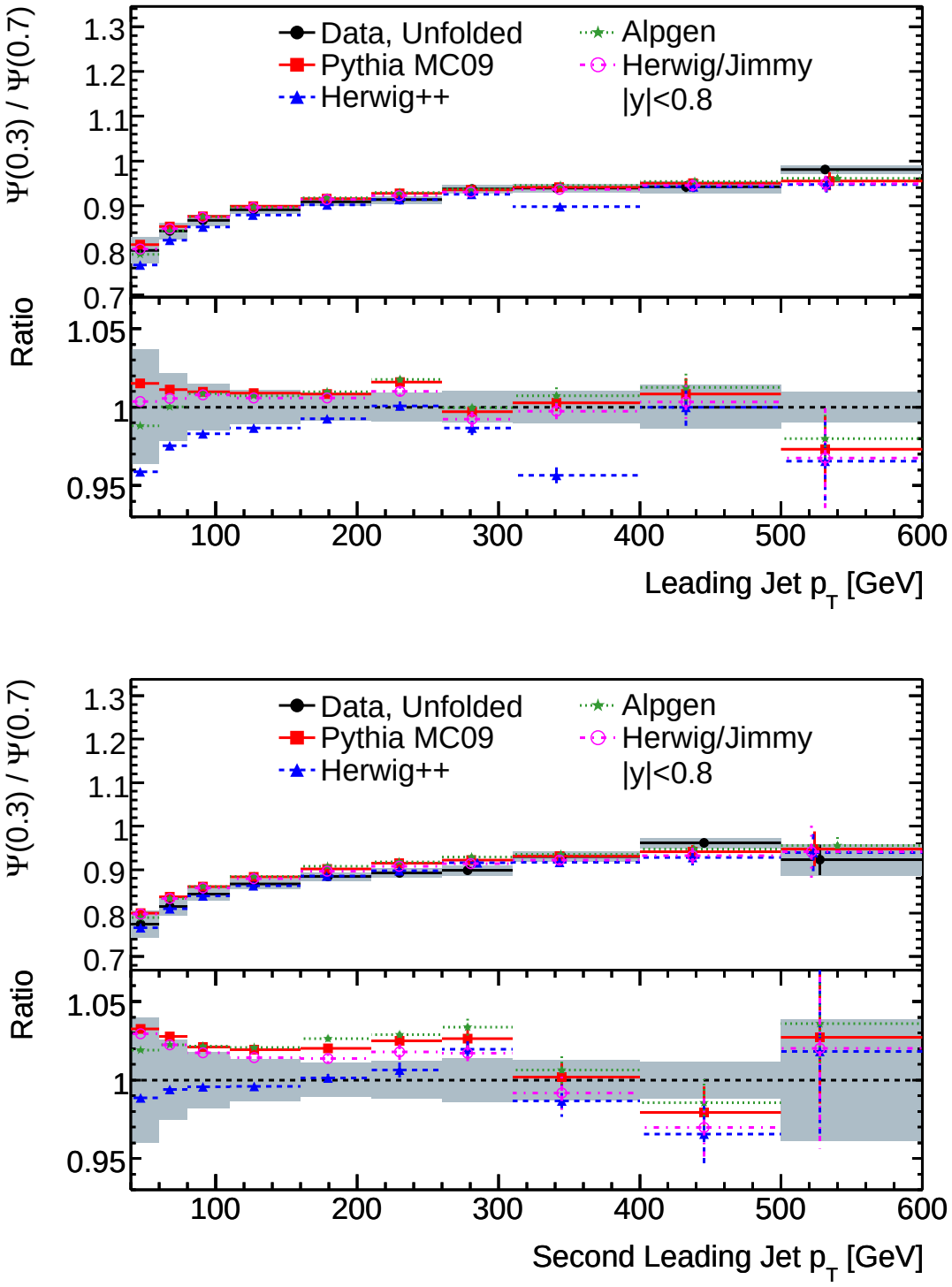


Figure 6.13: Jet shapes for the leading (top) and second leading (bottom) jet after unfolding, compared to those in several different Monte Carlo generators.

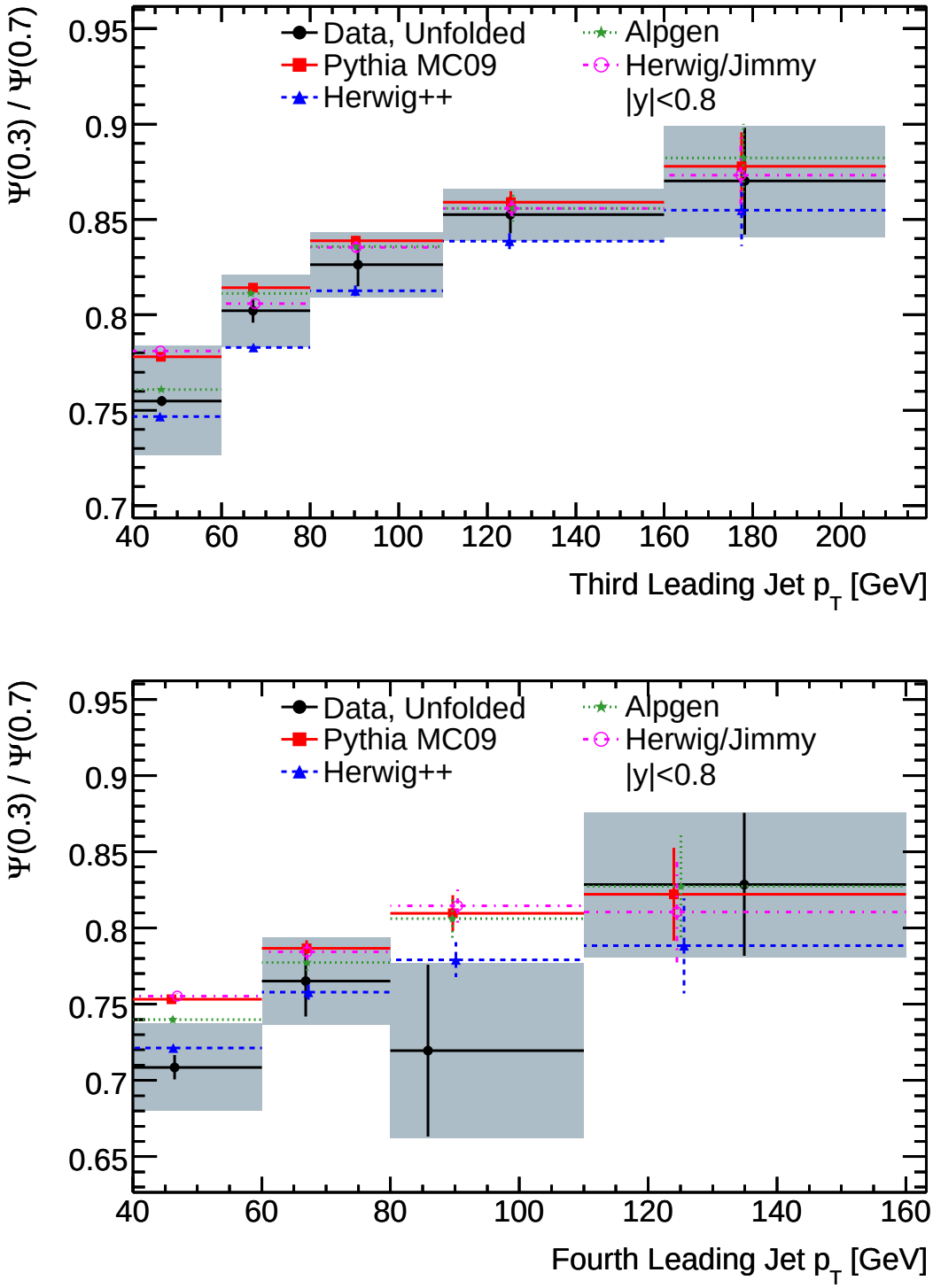


Figure 6.14: Jet shapes for the third (top) and fourth leading (bottom) jet after unfolding, compared to those in several different Monte Carlo generators.

6.3 Impact on Future Measurements and Searches

The measurement of jet shapes and subsequent evaluation and tuning of Monte Carlo impacts other physics measurements in two significant ways. First, many measurements require that a connection be made from jets measured in the calorimeter to partons coming from the hard scatter of the event. The shape of the core of the jet is primarily dictated by hard radiation and hadronization processes, whereas the shape at the edge of the jet and just beyond the jet cone is primarily dictated by the underlying event. Ensuring that all three of these contributions are correctly modeled is critical to establishing a link back to partons. Second, with the current jet calibration scheme in ATLAS there is a significant dependence of jet response on the flavor of the leading parton in the jet. This results in an inflated jet energy resolution and an additional term in the jet energy scale systematic uncertainty for any measurement using an event selection different from the one with which the jet energy scale was derived. The power of jet shapes as a discriminant provides some hope for substantially reducing this sample dependence and making the jet energy scale more universal.

When reconstructing heavy decays (e.g. $t\bar{t} \to b\bar{b}q\bar{q}\ell\nu$, $H \to b\bar{b}$, or $X \to q\bar{q}$), a connection must be made between the energy of the reconstructed jet and that of the parton from the decay. This connection requires understanding the underlying event contribution to jets as well as the losses due to hadronization and hard radiation which might occur. Examining the jet shapes outside of the jet, as discussed in Chapter 6, assists with the underlying event correction in two ways. The underlying event contributes more substantially to forward jets than to central jets. Thus, disagreements in jet shapes away from the jet core that strongly vary with rapidity are a good indication of a poor modeling of the underlying event. From this study, Herwig++ appears to include too large a contribution from the underlying event, whereas most of the Pythia tunes include too little activity (see, for example, Figures 6.3 and 6.4). These observations, combined with the standard "away-region" underlying event analysis (see, for example, [194]) can be used to test the assumption that underlying event activity is uncorrelated with the jet direction, even in a system with significant color flow. The study presented in Section 5.6.5 was the first in ATLAS to demonstrate that the contributions from pile-up events could be directly subtracted from a jet with surprising accuracy. This was a necessary first step towards demonstrating that

the underlying event is sufficiently uncorrelated to be removed in the same way.

The connection to partons also requires understanding "out of cone" showering, where through either hard radiation or soft hadronization processes some energy is lost between the initial parton and the final jet. A correction for out of cone particles (i.e. those lost due to hadronization and showering in the detector) must be taken from the Monte Carlo simulation. Although some information can be gleaned from single particle response studies (e.g. [195, 196]), the energy regime covered by such studies is well below that of the core of a high-p_T jet. Such a correction can *only* be trusted if jet shapes in the data are well-described by the Monte Carlo. Additionally, the evolution of $\Psi(0.3)/\Psi(0.7)$ with jet p_T is mainly controlled by the rate of hard radiation from the leading parton in the jet. Ensuring that the hard radiation is well-described is important for correcting jets to the parton level. This hard-radiation also influences the number of close-by jets in a many-jet final state. In order to correctly translate the number of jets in an event measured in the detector to the number of partons in the hard scatter, the hard radiation from the leading partons must be well-modeled. Because the rate is considerably higher for hard radiation from a gluon (or a splitting $g \to q\bar{q}$) than for hard radiation from a quark (e.g. the DGLAP equations predict a factor of two higher probability for $g \to gg$ than for $q \to qg$), the evolution of this shape is, to some degree, influenced by the quark-gluon composition of the sample.

Corrections for these effects are already being used by ATLAS in Standard Model measurements and in searches for new physics. For example, ATLAS has begun searching for di-jet resonances in 7 TeV proton-proton collisions [197, 198]. As long as limits are being set, these corrections enter only as a systematic uncertainty on the scale of the reconstructed resonance and result in a somewhat-reduced search sensitivity. However, if any resonance is found, the corrections will be crucial for measuring the mass of the newly-found particle(s). In the same way, top-quark reconstruction in ATLAS currently proceeds by taking all jet-to-hadron corrections directly from the Monte Carlo (see, for example, [199]). Only when the Monte Carlo correctly describes the underlying event, hadronization, and hard radiation effects present in the data will such a measurement yield an unbiased top-quark mass. The results shown in Chapter 6 imply that the Pythia tunes DW and Perugia2010 are preferable to the standard ATLAS Monte Carlo tune or to any of the Herwig variants considered for deriving these corrections.

An additional significant issue is the flavor-dependence of the jet energy scale calibration

and the possibility of alleviating this problem with a quark-/gluon-jet shape discriminant. The current jet calibration uses a simple p_T- and η-dependent scaling factor to correct jets for the non-compensation of the calorimeter (see Section 5.2). However, because of the different showering and fragmentation, quark-jets and gluon-jets have significantly different average energy responses in the calorimeter. The difference in response between quark and gluon-jets is show in Figure 6.15 for Anti-k_t jets of size 0.4 and 0.6 [200, 201]. There is a difference of up to 7% at low p_T, falling to ~2% at high p_T[1]. Both cone sizes are included here to emphasize that this is not simply a result of a too-small or too-large jet size. Because it affects both large and small cones, virtually all measurements will be impacted by the flavor-dependence of the jet response. This flavor-dependence is largest when using a simple scale for jet calibration. More advanced jet calibration techniques that take into account the properties of the jet, including the jet's width, the depth of the shower in the calorimeter, and the energy density in the core of the jet, may reduce this dependence [201]. The size of the reduction will depend on the final implementation of the jet calibration that is applied. However, no such calibration has yet been validated for use in ATLAS.

The jet calibration used for all ATLAS studies is based on the event selection of the inclusive jet cross-section measurement [170, 172, 173], which is the same event selection as was used for this analysis. Importantly, the calibration was done for a specific combination of quark-jets and gluon-jets. Studies with different event topologies must include the flavor-dependent response as a part of the jet energy scale systematic uncertainty. For example, at low-p_T, the inclusive jet spectrum is dominated by gluon-jets. The top-quark mass, however, when measured in either fully-hadronic or semi-leptonic final states (i.e. final states with one or both W-bosons decaying into $q\bar{q}$), includes a number of low-p_T quark-jets. The top-quark mass would, therefore, be measured incorrectly as a result of the flavor dependence of the jet response. Moreover, the difference in response may lead to lost $t\bar{t}$ events if the event selection cuts applied are sufficiently tight. Measurements using sub-samples are affected by this uncertainty as well. The measurement of the number of jets per event, for example, uses the same event selection as the inclusive jet cross-section measurement. However, the sample is further sub-divided according to the jet multiplicity of the events. The flavor content of those sub-samples, particularly the high multiplicity events, is not well-known.

[1] The non-closure of response (i.e. the fact that the average jet response is not identically one) is included in the jet energy scale systematic uncertainty and is, in part, a result of the calibration based on jet energy when the figure shows response as a function of jet transverse momentum.

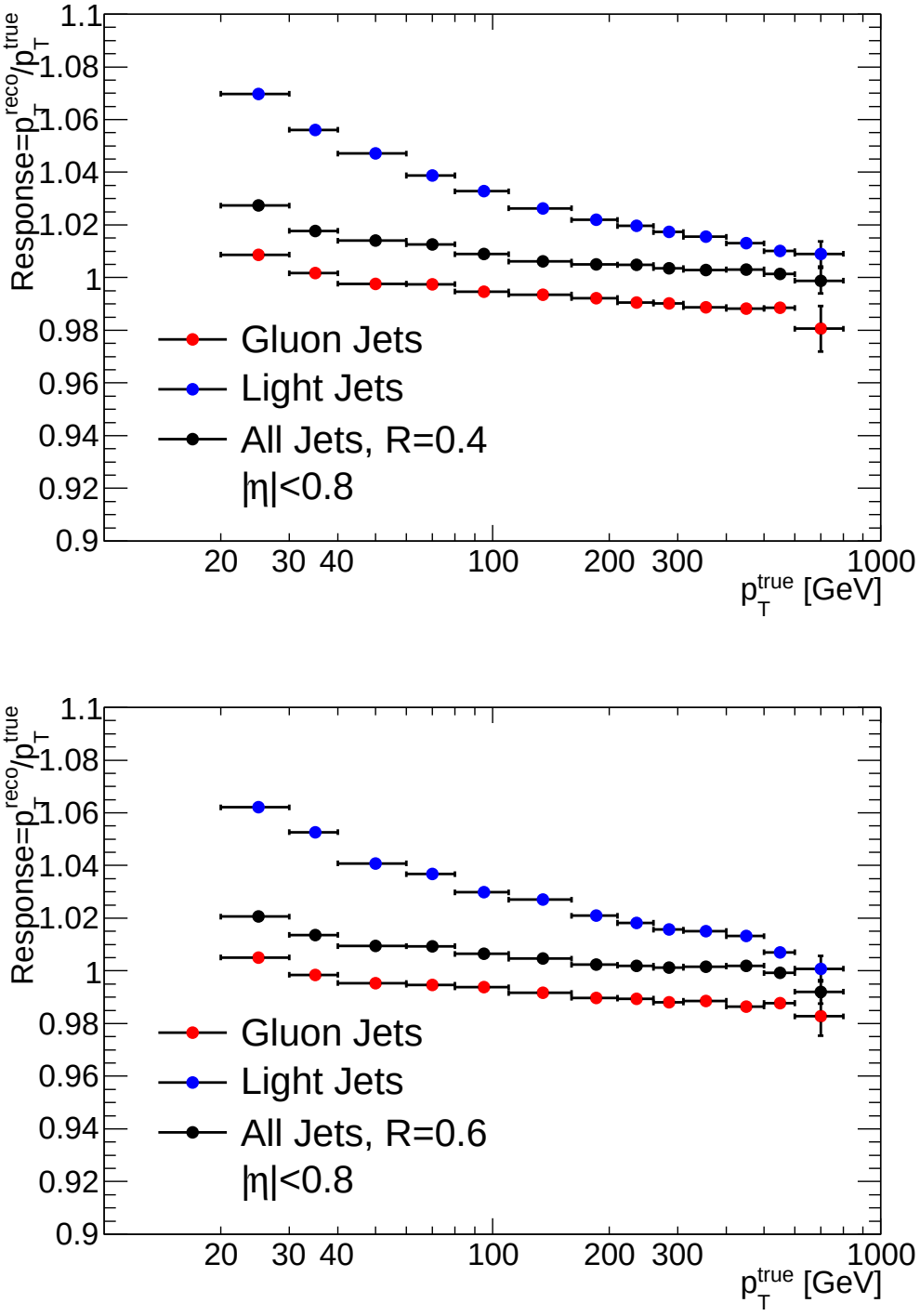


Figure 6.15: The response of quark-jets and gluon-jets after calibration to the hadronic scale with a p_T- and η-dependent scaling factor for Anti-k_t jets with a radius of 0.4 (top) and 0.6 (bottom). A difference of up to 7% is visible at low p_T, falling to ~2% at high p_T.

Thus, the bias in response must be taken as a systematic uncertainty on the measurement. Without any work to reduce the flavor-dependence of the jet energy scale, this systematic uncertainty would eventually dominate the total uncertainty on the measurement.

This problem is not alleviated by *in situ* jet calibration techniques. One common calibration scheme is to require a transverse momentum balance between photons, objects that are very well measured by the calorimeter, and jets, objects which are much more difficult to correctly calibrate due to the non-compensation of the calorimeter [202]. However, the balancing jets in these events are predominantly quark-jets. Thus, the calibration derived for these events is not applicable to samples with moderate or large gluon-jet content. Particularly for the energy range that is best calibrated using this technique, the inclusive jet sample is dominated by gluon-jets, leading to a near-maximal calibration bias.

For current measurements using different event selections, this bias must be included as an additional term in the jet energy scale uncertainty. As the jet energy scale uncertainty is ∼7-9% without the additional flavor-dependent term, it increases the total uncertainty by ∼20-40%. However, with the significantly reduced jet energy scale uncertainty that is expected to follow on the improved detector calibration at the end of the 2010 data taking period, the flavor-dependent term of the jet energy scale uncertainty will be by far the dominant part.

In order to reduce the flavor-dependence of the jet energy scale, the width variables developed in Section 6.1 can be used to provide a parameterized *post-hoc* jet correction. Applying such corrections has been studied previously in ATLAS [86, 87]. The result of the correction is two-fold. First, the jet energy response dependence on the jet flavor is reduced. Equivalently, the jet energy scale systematic uncertainty for any study using events different from those of the inclusive jet cross-section measurement is reduced. Second, by bringing the response of different classes of jets closer to that of the inclusive sample, the jet p_T resolution is reduced.

In order to evaluate the significance of this bias and the importance of applying some correction in a physics measurement, the reconstruction of the top-quark mass was examined. $t\bar{t}$ events were produced using the MC@NLO event generator [203] interfaced to Herwig and Jimmy for hadronization and underlying event modeling [57, 163]. In order to fully isolate the effect of flavor-dependent jet mis-calibration, no detector simulation was applied. One of the top-quarks was required to decay leptonically ($t \to b\ell\nu$, where ℓ is either

an electron or muon), and the other was required to decay hadronically ($t \to bq\bar{q}$). This decay topology is the most interesting for examining the jet energy scale dependence on flavor, since it requires the reconstruction of exactly two quark-jets from the hadronically decaying W. Mis-calibrating light quark-jets can lead to either a poor mass reconstruction or missing a top-candidate altogether.

At the Tevatron, the total cross section for top-quark production was almost 25 times lower than it is in 7 TeV collisions at the LHC [204]. The higher center-of-mass energy of the LHC also results in many more events with several jets in addition to those from the $t\bar{t}$ system. These additional jets, which are typically from gluon initial- or final-state radiation or from secondary parton-parton interactions, result in a higher combinatoric background at the LHC than was present at the Tevatron. Thus, the problems resulting from mis-calibrating only quark-jets are more significant at the LHC and must be treated with additional care.

Selection criteria were defined in order to closely follow the standard ATLAS top-quark analyses [87, 199]. Leptons were required to have $p_T > 20$ GeV and $|\eta| < 2.5$. Jets were tagged as "b-jets" if they included a hard b-quark in the parton shower that pointed towards the jet. The flavor tagging of light-flavor (light quark and gluon) jets proceeded as in Section 6.1. All jets were required to have $p_T > 20$ GeVand $|\eta| < 4.5$. The b-jets were additionally required to have $|\eta| < 2.5$ in order to account for the end of the acceptance of both the inner detector and the muon spectrometer, one of which is necessary to experimentally tag reconstructed b-jets. If a jet was within 0.2 in $\Delta R \equiv \sqrt{\Delta\eta^2 + \Delta\phi^2}$ of a lepton and had less than twice the lepton's p_T, the jet was rejected. If a lepton was between 0.2 and 0.4 from a jet in ΔR or was within 0.2 in ΔR and had a p_T below half of the jet's p_T, the lepton was rejected. The p_T-dependent overlap removal is necessary in order to avoid counting leptons or rejecting jets from leptonic b-quark cascades. Roughly 25% of b-jets decay leptonically, and for these decays the cut is highly efficient.

For each lepton in the event, a W-boson was reconstructed by assuming that the missing transverse energy in the event corresponded the high-energy neutrino from the leptonic W-decay. The rapidity of the missing transverse energy was chosen to give the correct W-mass. These leptonic W-bosons were combined with the b-jets in the event to find combinations within 40 GeV of the top-quark mass. Hadronic W-bosons were reconstructed from pairs of non-b-tagged jets in the event and were required to be within 20 GeV of the W-mass.

The b-jets not matched to the leptonic top-quark decay were combined with the hadronic W-candidate to form a hadronic top-quark candidate. No attempt was made to correct for out-of-cone showering or the addition of underlying event to the jets (i.e. to fully correct the jets to the parton level). For that reason, the exact value of the reconstructed top-quark mass is not relevant, but the shift in reconstructed mass that results from a flavor-dependent jet mis-calibration is.

Two scenarios were considered. First, all the object reconstruction in the event was done using the true p_T of the jets in the event ("ideal reconstruction"). Then, all jet momenta were shifted according to the bias shown in Figure 6.15 ("mis-calibrated reconstruction"). In this case, jets tagged as gluon-jets were shifted by the bias shown in gluon-jet response, and quark-jets were shifted by the (much smaller) bias shown in quark-jet response. Because they have shapes similar to gluon-jets [111], b-jets were shifted by the smaller quark-jet bias. In order to be self-consistent, the entire event reconstruction was done after applying the shift (i.e. different combinations of jets were allowed to result in reconstructed top-quarks before and after the momentum shift).

Figure 6.16 shows the reconstructed hadronic W-boson mass in ideal reconstruction and with the flavor bias in jet response. Over the fairly large combinatoric background the W-boson mass peak is clearly distinguishable. However, with the mis-calibration of light quark-jets the position of the peak is shifted by ∼5%. For a tight cut on the hadronically reconstructed W-boson mass, this also implies a significant drop in top-quark reconstruction efficiency. An average correction for this shift may be derived specially for this measurement, but if the jet shapes are not correctly described by the Monte Carlo then the correction that is derived would likely be inaccurate.

Figure 6.17 shows the reconstructed hadronic top-quark mass in ideal reconstruction and with the flavor bias in jet response. The top-quark mass peak is again clearly distinguishable over the combinatoric background. The flavor-dependent calibration bias induces a ∼3% (4.6 GeV) shift in the reconstructed top-mass, a shift about four times larger than the current world-average top-quark mass uncertainty [205]. For these figures, Anti-k_t jets of size 0.6 were used. Comparable biases were observed in Anti-k_t jets with sizes of 0.4 and 0.6.

For a measurement of a well-known final state like $t\bar{t} \rightarrow b\bar{b}q\bar{q}\ell\nu$, it may be possible to apply a first correction for the flavor mis-calibration directly from Monte Carlo. The

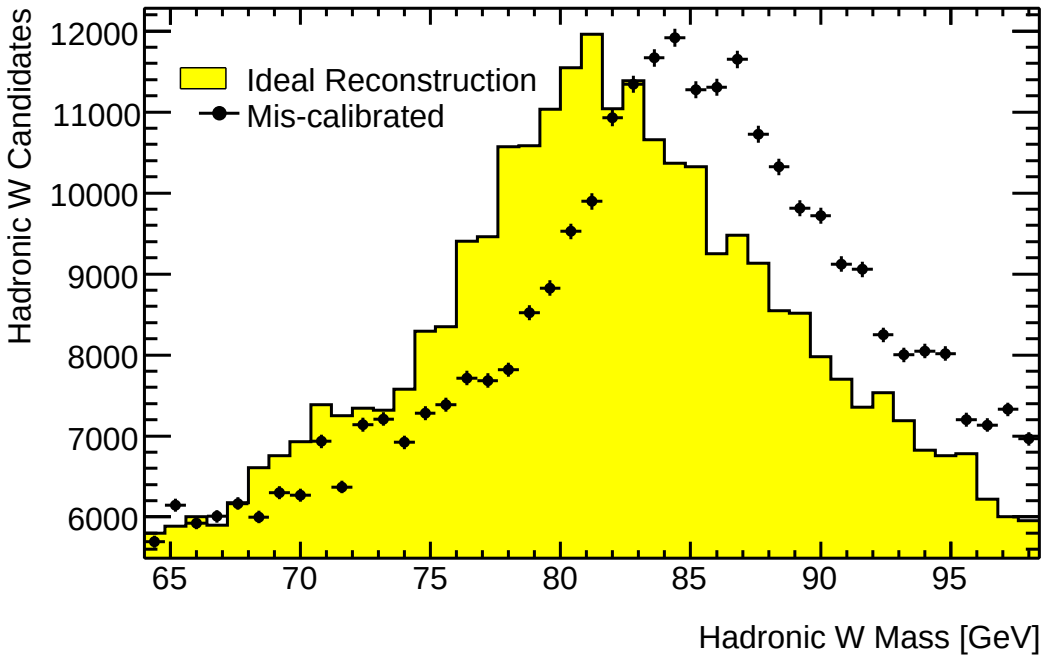


Figure 6.16: The reconstructed hadronic W mass with and without a flavor-dependent jet calibration bias. A ∼5% shift to higher mass is visible over the combinatoric background.

quark/gluon mixture of the final state is well-known for events without additional jets from radiation. Provided the jet shapes in data are sufficiently well-described by the Monte Carlo (i.e. if the underlying event and hadronization effects are well-described), the inverse of the bias shown in Figure 6.15 can be used as a correction to the jet p_T. However, this approach is only possible in final states where the quark/gluon mixture is well known. For new physics scenarios, the mixture will be ambiguous, and such a correction will not be possible. Different scenarios of Supersymmetry, for example, could result in hard jets with either leading quarks or leading gluons [206]. A flavor-dependent mis-calibration may result in both an artificial broadening of a new-physics resonance and the incorrect measurement of a new mass peak or decay edge.

There are many secondary benefits of the jet shape measurements that will improve the uncertainties on future measurements and the reach of future new physics searches. For example, one of the most common ways to reduce the multi-jet background in new physics searches is by requiring an isolated lepton. The rejection power of a set of electron identification criteria, however, is heavily dependent on the shape of the jets that make

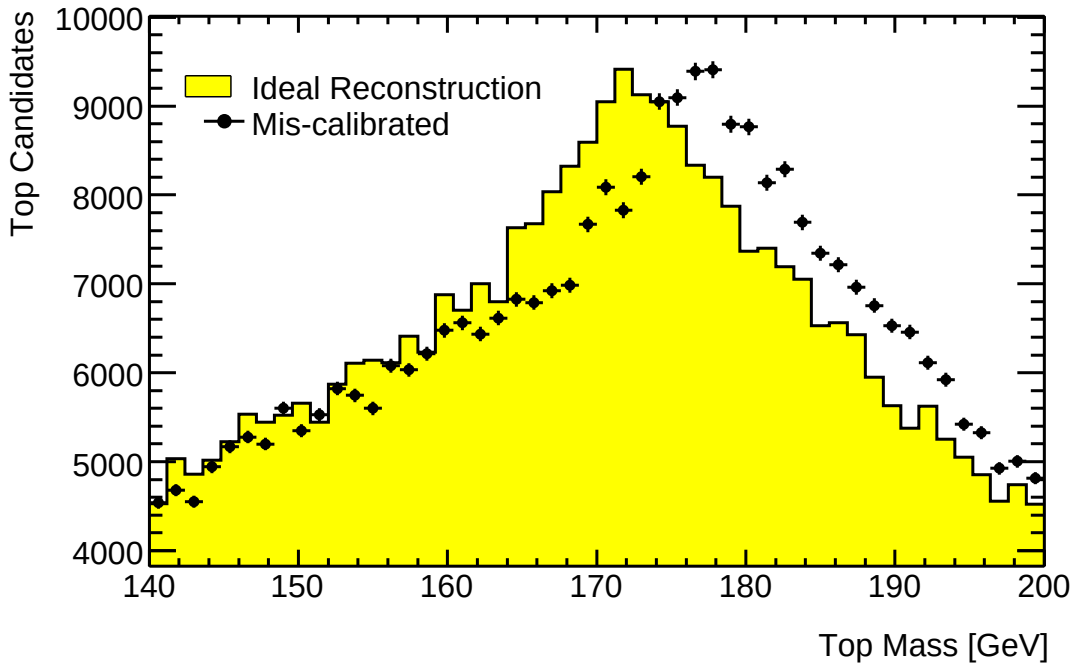


Figure 6.17: The reconstructed hadronic top-quark mass with and without a flavor-dependent jet calibration bias. A ∼3% shift to higher mass is visible over the combinatoric background.

up the QCD background. Because the jet production cross section is several orders of magnitude higher than the electron production cross section, even a small change in the jet rejection power can introduce a large QCD background in a search. Without a good understanding of jet shapes in data, the systematic uncertainty on this rejection power remains large.

The inclusive jet spectrum measurements [172, 173] soon will be used to constrain parton distribution functions (PDFs) at high-Q^2 and to measure the strong coupling constant, α_s. Because the jet calibration is performed using the same sample as the inclusive jet cross-section measurement, there is no flavor-dependent term in the jet energy scale uncertainty for this particular measurement. However, the flavor-dependence of the jet calibration introduces an ambiguity in theoretical interpretation that leads to an inflated uncertainty on the constraint of the PDFs and the measurement of α_s. With the bias shown in Figure 6.15, for example, a larger light-quark fraction could produce the same effect as a harder jet p_T spectrum (i.e. a larger α_s). As the PDF uncertainty often dominates the measurement of α_s [207], correcting this bias is critical to producing a competitive result.

The potential benefits from the correct measurement of jet shapes and the resulting adjustment of Monte Carlo to correctly describe the data (or, conversely, the potential detriment of an incorrect description of jet shapes in the Monte Carlo) are extensive. The most common interpretation of a jet as representing a unique parton from the hard scattering, a connection which is theoretically difficult to begin with, is complicated further by corrections from the underlying event, hadronization, and hard radiation from the leading parton. Without a correctly tuned model of the behavior of QCD in these non-perturbative and difficult to calculate regimes, the experimental corrections normally applied to translate jet measurements to a partonic interpretation are invalid. Measurements of Standard Model processes as well as searches for new physics are improved by a complete analysis of jet shapes.

Chapter 7 Conclusions and Future Prospects

We have presented an analysis of jet shapes in an inclusive jet sample from proton-proton collisions at $\sqrt{s}$ =7 TeV. The jet shapes $\rho^a(r)/\Psi(0.7)$ and $\Psi(0.3)/\Psi(0.7)$ were compared to predictions from several different Monte Carlo generators for jets with $40 < p_T < 600$ GeV and $|y| < 2.8$. The agreement between the data and Monte Carlo was found to be within the systematic uncertainties in most regions of phase space. However, there is some significant disagreement, particularly at low p_T and high rapidity. The new Monte Carlo generator HERWIG++ shows some promise in being able to describe jet shapes near the center of the jet but shows significant disagreement far from the jet core. Some of this discrepancy is likely due to a poor modeling of the underlying event.

Many of these basic jet shapes are critical for evaluating the jet energy calibration applied by ATLAS. It is necessary to have a correct description of the fraction of the shower beyond the calorimeter region included by the jet algorithm. As high-p_T jets from QCD are significant backgrounds to new physics searches, a detailed calibration such as the one envisioned by ATLAS will need to incorporate these effects. The disagreement observed between data and most Monte Carlo predictions may soon be resolved by a fresh tune of the generator parameters. The good agreement of the Perugia2010 tune with the available data indicates the importance of including these jet shape measurements in the tuning input data set.

Two new jet shape variables were presented which show significantly improved power to distinguish between quark-jets and gluon-jets relative to the historical shape variable definitions. For forward jets, there is significant disagreement between the data and a sample of only quark-jets in PYTHIA. As the shapes of the inclusive samples do not agree, of course, no conclusions can be drawn about the quark content of the inclusive jet sample. With further tuning and an improved systematic uncertainty on the unfolding, it might be possible to make stronger statements about the agreement of the Monte Carlo generators with different flavor jets, and perhaps even about the flavor content of a sample of events. Of course, in a proton-proton collider it is not straightforward to construct a sample of gluon-jets or quark-jets with significant purity. However, by exploiting event topologies like

γ/jet or the evolution of the quark fraction with transverse momentum and rapidity, it may be possible to create samples with sufficiently different quark/gluon composition to test a likelihood variable.

As the ATLAS Collaboration continues to collect data, improve its understanding of the detector, and improve the agreement of the Monte Carlo with the data, several more studies will be possible which were not included in this early work. Once the more advanced jet calibrations have been validated, it should be possible to construct jet shapes from both calibrated and uncalibrated constituents. As long as a simple bin-by-bin unfolding is applied, such a correction may make it possible to reduce the unfolding systematic uncertainty. With higher statistics, provided the effect of pile-up is kept under control, it should be possible to more precisely measure sub-leading jet samples and make a stronger statement about the importance of the higher-order matrix elements. Finally, once a large sample of heavy-flavor jets has been collected it should be possible to measure the jet shapes separately in the inclusive sample and the heavy flavor sample, following [110]. Analyzing jet shapes in these corners of phase space will provide a more complete test of the current models.

Two additional studies will be considered briefly here. First, an alternative unfolding technique will be discussed, which could reduce the systematic uncertainty due to the unfolding in the future. Second, some newer jet shape variables will be examined. These shapes are particularly interesting when searching for new physics as discriminants between QCD jet background and signal events. With the currently available statistics, it is only possible to inspect the distributions in a much lower p_T range than the one in which they are most interesting.

7.1 Unfolding Techniques

The largest systematic uncertainty in this analysis was from the unfolding. The uncertainty may be reduced by new generator tunes which better describe the data. In order to minimize the systematic, a new technique for unfolding was developed which differs fundamentally from other more advanced unfolding methods.

The most common advanced unfolding techniques involve the construction of a "response matrix" in the observable [175, 208]. In the case of radial jet shapes, this response matrix

would be unique for each bin of p_T and rapidity, with one column and one row for each radial bin. The entries in the matrix would describe the migration of energy from one annulus into another. For a hadron-level radial jet shape $\vec{h}$ (a vector, with one element for each radial bin) and a calorimeter-level radial jet shape $\vec{r}$, the unfolding would then be written as:

$$\vec{h} = \mathbf{C}^{-1} \times \vec{r} \tag{7.1}$$

where $\mathbf{C}^{-1}$ is the inverted response matrix. In the case of jet shapes, such a derivation relies on the detailed tracing of particles from the hard scatter to calorimeter deposition. Effects like calorimeter thresholds and noise make the assignment ambiguous at best. Further, the particle spectra within the jets must be similar between data and Monte Carlo for the unfolding to be valid - or additional dimensions must be added to the response matrix to account for possible variations. Even if such a response matrix can be constructed, it must then be inverted. The inversion of these matrices, which may not be invertible or may suffer from non-physical results, has created a small industry of work [208]. Some of the proposed alternatives, like a Bayesian iterative bin-by-bin unfolding [209], still involve the creation of a "smearing matrix." While these methods avoid the problem of matrix inversion, they still suffer from the difficulty, at least in the case of jet shapes, of constructing a sensible matrix.

Off-diagonal elements in the correction, however, are important during unfolding. These elements take into account flow from one bin to another. For example, jet energy carried by soft particles will be likely to move bins as the particles bend in the magnetic field. More soft particles in the center of the jet result in not only a lower central jet shape measurement, but also a higher measurement at large radius.

There is an alternative approach to constructing the unfolding matrix that does not rely on the detailed tracing of particles in simulation but still takes into account flow from one part of the jet to another. Conceptually, it relies on a simple formulation that can be thought of as a mathematical attempt to minimize the unfolding bias. Practically, it can be applied to any unfolding problem, requires no further information than the final distributions of the observables, and avoids any matrix inversion. For every Monte Carlo sample i, one can define a $\vec{r}_i$ and an $\vec{h}_i$ (i.e. a separate vector for PYTHIA MC09, PYTHIA

Perugia2010, HERWIG, etc). Then, one can attempt to find the matrix $\mathbf{C}$ such that:

$$s = \sum_i w_i \times \left| \mathbf{C} \times \vec{r_i} - \vec{h} \right|^2 \tag{7.2}$$

is a minimum, where w_i is a weight for each sample, and || denotes the magnitude of the vector. This is simply a formulation of a least squares problem, which is a type of problem that has been studied (and solved) since the time of Gauss and Legendre. In approaching the problem in this manner, one is able to achieve an unfolding with off-diagonal elements (i.e. taking into account bin-to-bin migrations) without ever constructing a response matrix in the traditional manner.

The unfolding matrix derived using this method is not biased by relying on only one sample, as the classical bin-by-bin unfolding method is. Instead, it finds the matrix that best removes the detector effects from the simulation and returns all of the various input samples to their "truth." As a result, there is no need for nature to be bracketed in any way by these samples. Rather, in order to minimize the unfolding bias one need only include more samples in the input set. In some sense, the different Monte Carlo samples (e.g. PYTHIA and HERWIG) that have been run through the same detector simulation act as different representations of nature. Uncertainties in the detector simulation itself can be fairly included as samples using a single Monte Carlo generator but a variant of the detector simulation. In this method, one is even free to include possibilities that seem unlikely (e.g. more extremal tunes or unrealistic variations in detector conditions) with a reduced weight if, for example, more appropriate samples are not available.

7.2 New Jet Shapes

Several new physics models involve quarks and gluons in the final states (see, for example, [87, 210]). However, the QCD jet background typically has a cross section many orders of magnitude higher than the production cross section for any new particles. In order to reduce the background, several new jet shapes variables have been proposed. One such variable, discussed in Section 2.2, is the "pull" vector [113]. This jet shape is particularly helpful in distinguishing QCD di-jet production from di-jets coming from the decay of some colorless particle. Pull relies on color connection in theory and relies on an experiment's

angular resolution and ability to measure the energy deposits near the edge of a jet. In principle, a jet from QCD will have a pull vector that points in the $\vec{\eta}$ direction (i.e. towards the beam pipe) because of color connections to the beam remnants. In the case of $Z \rightarrow q\bar{q}$ (also $W \rightarrow qq$' and $H \rightarrow q\bar{q}$), the pull vector of the first quark jet will point towards the second (roughly in the $\vec{\phi}$ direction) and vice-versa because of color connections between the two quark jets.

As this jet shape has not been evaluated in any previous experiment, it is important to consider the experimental effects that may help or harm its use. Any differences in the detector between η and ϕ resolution could reduce the possible discrimination power of the variable. Because of experimental effects, the jets in the ATLAS calorimeter are not, in fact, symmetric in η and ϕ. For understanding the issues that produce this asymmetry, it is simpler to examine the second p_T moments of the jets. Figure 7.1 shows the second moments in η and ϕ as a function of jet p_T constructed at the hadron level in Monte Carlo, compared to the same moments constructed from calorimeter quantities (either towers or clusters, see Section 5.2). In all cases, the second moment decreases with rising jet p_T, indicating that the jets are more collimated at higher p_T. The rejection of low p_T particles near the edge of the jet (due to material effects, noise suppression thresholds, and the magnetic field) results in some of the observed reduction in the second moment. For measuring soft radiation near the edges of jets, as the pull variable attempts to do, these losses can be important.

Another key feature of the second moments distributions is the asymmetry between η and ϕ. At the hadron level, the second moment is symmetric between η and ϕ. The magnetic field will smear the jet in ϕ, creating a wider second moment in ϕ than in η, just as is observed. That jets in a real calorimeter tend to be broader in ϕ than in η means that the pull vector of the QCD di-jet background events may tend to point randomly, rather than being predominantly pointed towards the beam pipe. As signal jets are meant to point in the $\vec{\phi}$ direction, this effect may harm the signal to background discrimination of the pull vector. Pile-up, wherein a single event has more than one proton-proton interaction, will also have a negative effect on the variable's discrimination power, since it will tend to randomly distort the pull vector for a jet. An additional cut on jet ellipticity or pull vector magnitude might ensure that the pull vector is less susceptible to energy fluctuations around the jet. Using tracks to create the pull vector may also reduce some of these difficulties, in particular by reducing the minimum p_T of the included particles and by allowing some

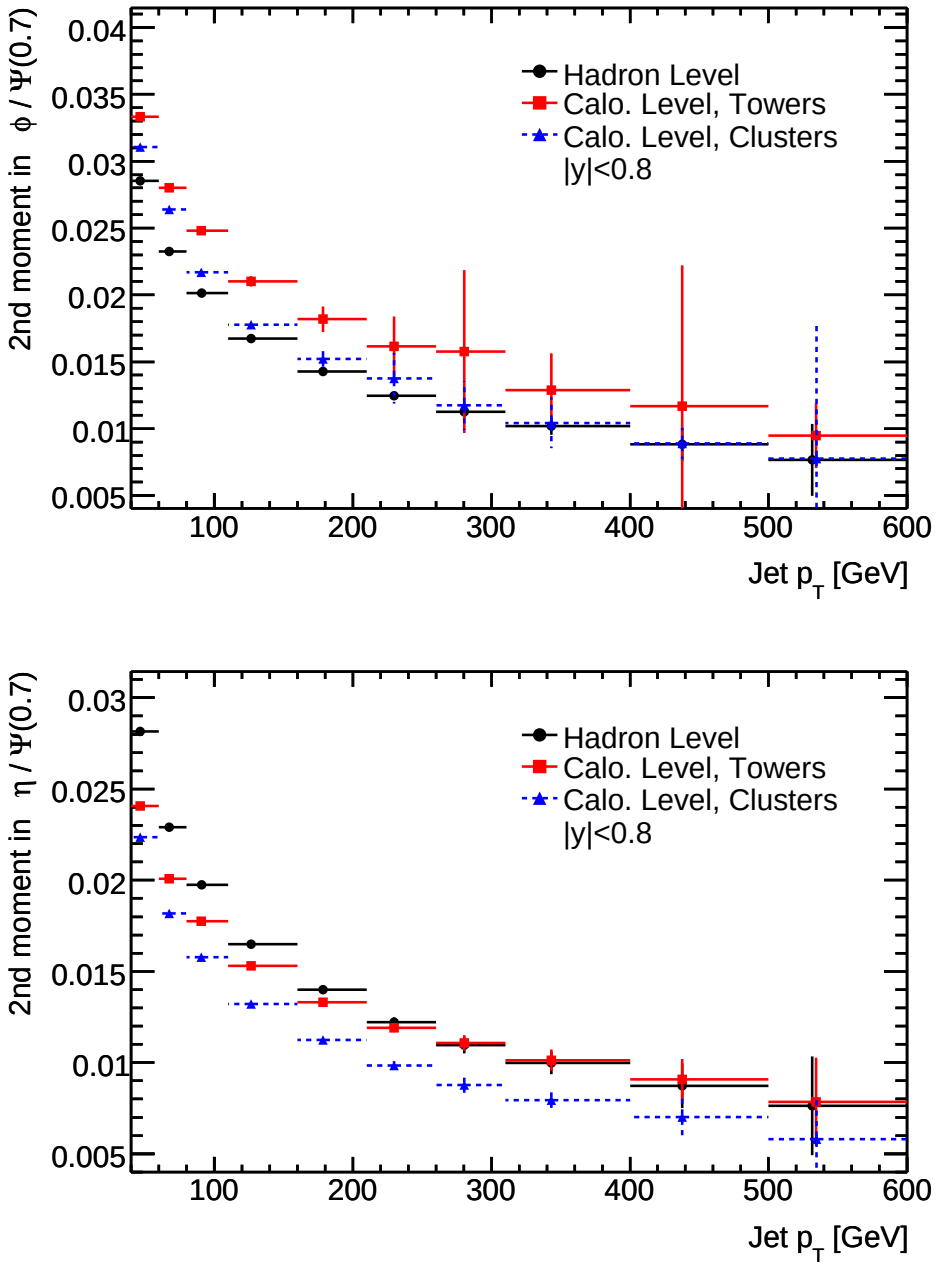


Figure 7.1: Jet second moments in ϕ (top) and η (bottom), built at the hadron level and at the calorimeter level using both towers and clusters, in Monte Carlo Simulation.

discrimination between particles from different proton-proton interactions. However, such a construction would introduce a dependence on the fraction of the particles that are charged, and so would have to be studied carefully.

There is an additional effect visible in the differences between the second p_T moments formed using hadrons and those formed using either towers or clusters. A cluster may include energy from several towers, and in forming jet shapes all the cluster's energy is placed at the energy-weighted centroid of that cluster. Towers, on the other hand, gather energy in a small, geometrically defined area. In some sense, towers force the energy distribution to spread. Thus, the second p_T moments for towers are consistently wider than those formed with clusters. The spread, however, is affected by the geometry of the cells within the tower. In the limit that a tower consists of a single cell, the energy of that tower will be placed, by definition, at the geometric center of the tower. In the limit that a tower consists of an infinite number of cells, the energy centroid of the tower is free to move, in principle, anywhere within the geometric limits of the tower. In the ATLAS calorimeter, many towers have finer granularity in η than in ϕ, allowing larger shifts of the energy centroid in η. From this effect, one expects towers to produce shapes closer to those of clusters in η than in ϕ. This is one possible explanation for the differences observed in Figure 7.1.

Pull will require some study in data in order to ensure that it is not susceptible to these experimental effects. Nonetheless, the pull vector for jets in the calorimeter can already be constructed and examined in Monte Carlo simulation. Figure 7.2 shows the pull vector angle for jets in a single p_T bin and two different rapidity bins. In these QCD di-jet events, there is, as expected, an excess of jets with pull vectors pointing towards the beam pipe (angles near zero and π). Moreover, one can see that for more forward jets the pull angle tends to point towards the closer beam pipe. Interpreted as an effect of color connections, this implies that jets are more likely to be color connected to the outgoing parton or beam remnant that is closest to them in phase space. In the data available, Figure 7.3 shows the agreement between data and Monte Carlo simulation. The agreement is good with all the available generators.

Color connection nicely explains several of the features observed in Figures 7.2 and 7.3. The asymmetry in the pull angle is enhanced in samples with the Pythia DW tune, which has maximal color connection [191]. The fraction of jets that color-reconnect to the beam remnant opposite to them in phase space (i.e. those that form a pull vector pointing to

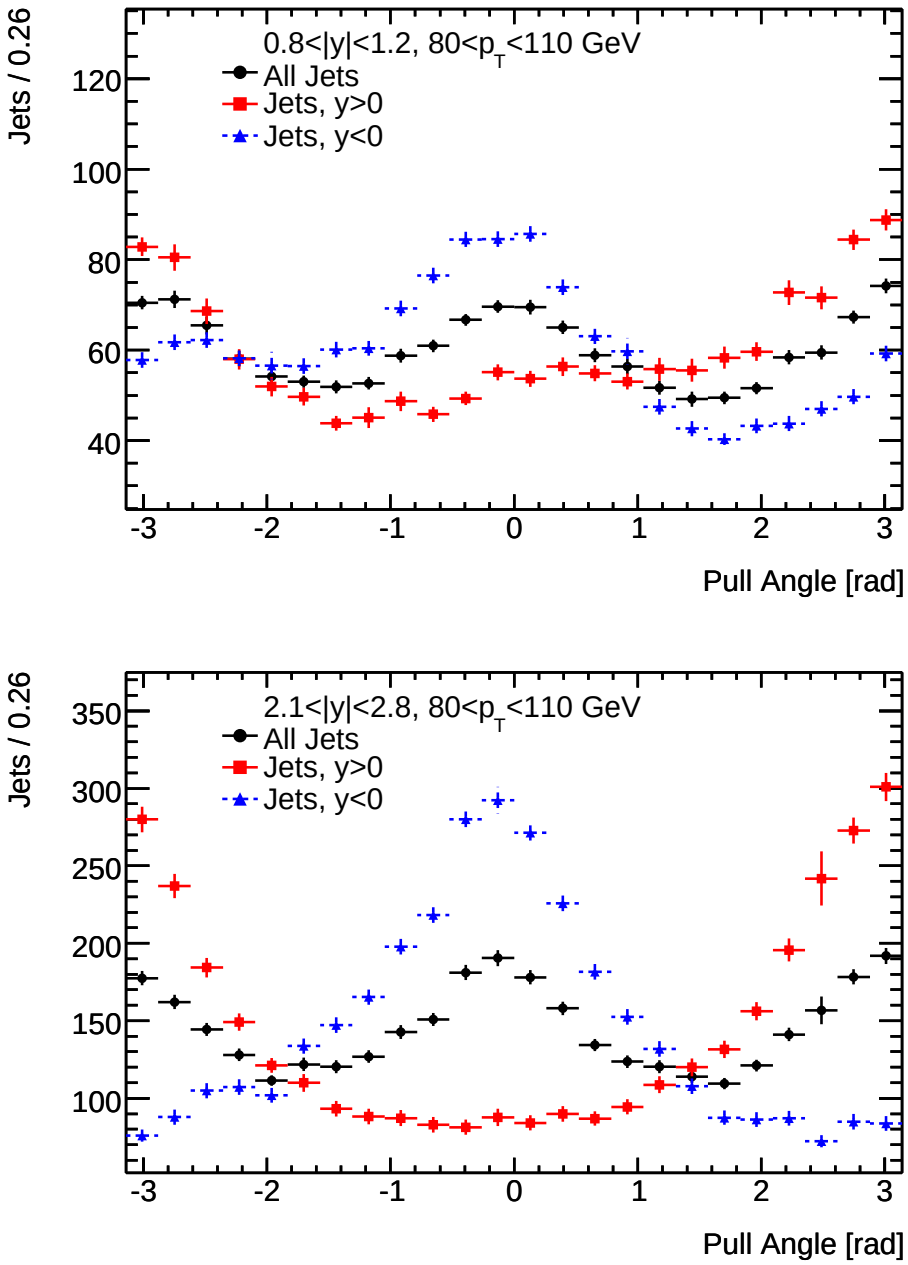


Figure 7.2: Pull vector angles for jets with moderate rapidity ($0.8 < |y| < 1.2$, top) and forward rapidity ($2.1 < |y| < 2.8$, bottom) in Monte Carlo simulation. An angle of zero indicates a pull vector that points towards the beam pipe in negative z.

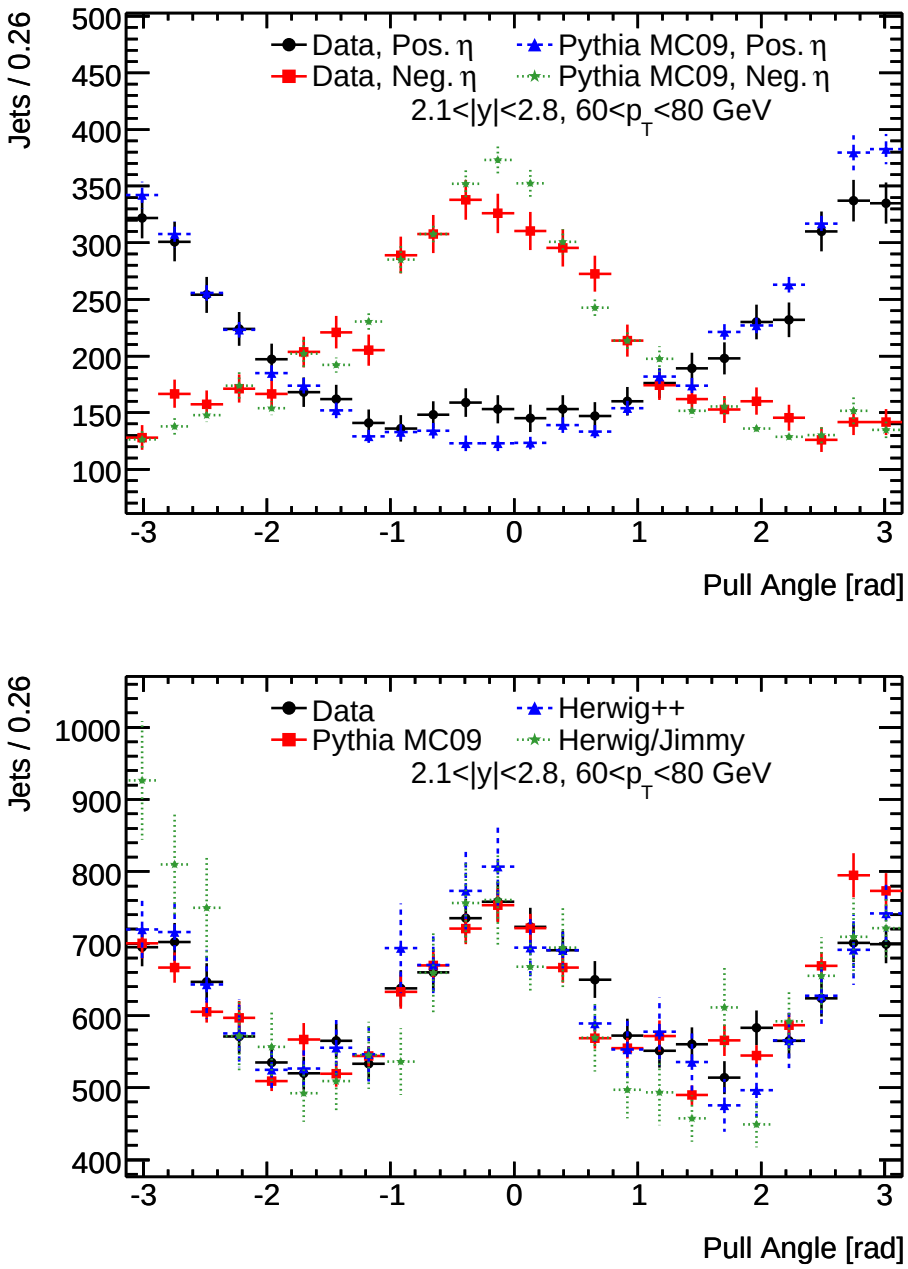


Figure 7.3: Pull vector angles for jets with forward rapidity ($2.1 < |y| < 2.8$) in data and Monte Carlo simulation. An angle of zero indicates a pull vector that points towards the beam pipe in negative z.

the further away beam pipe) should be about 10% [113]. This is consistent with what is observed in the sample of forward jets in Figure 7.2 and further suggests that the observed effect is a result of color connection in the QCD events. The difference between forward jets and central jets at constant p_T may also be explained by the higher fraction of gluon jets in the central region. Because gluon scattering may have more complicated color connections than the quark scattering, the color connection to the beam remnant may be suppressed. Additionally, because gluon jets tend to be wider than quark jets, they may be more susceptible to the threshold effects and distortions described above. At high p_T in both the signal and the background, quark jets are more common than gluon jets. For some searches, in particular searches for the Higgs boson decaying to two b-quarks, the signal and background are almost entirely di-quark-jet events. Thus, if the fraction of gluon jets is the explanation for these differences, this bodes well for the signal to background discrimination of the pull angle.

Measurements involving color connection will provide an opportunity to test differences between the available Monte Carlo programs as well. PYTHIA, with its integrated underlying event model, allows color re-connection between partons from the hard scatter and those from the underlying event. However, in part because the underlying event is controlled by an independent program, HERWIG does not allow such color re-connection schemes. Currently HERWIG++ does not allow re-connection to the underlying event, but it may in the future. Such re-connection would dilute any signal and may be important for understanding the pull vector's discrimination power. At the very least, the pull vector provides one test for the importance of correctly modeling color re-connection.

The ATLAS detector, despite its different calorimeter granularities in η and ϕ, and despite the magnetic field that spreads jets in ϕ before they reach the calorimeter, will apparently be able to resolve some color connection effects in QCD di-jet events. With sufficient data, it will, of course, be necessary to understand how these experimental effects might affect signal and background discrimination. More advanced jet shapes and studies, must wait for the Monte Carlo generators to be tuned to correctly model the most basic jet shapes. However, it is already possible to begin studying these shapes in preparation for the search for new physics.

Appendix A

This appendix contains the comparison plots from Chapter 5 and additional plots from Chapter 6. For all Monte Carlo simulation figures, the ratios are defined with respect to the hadron-level observable. For all figures including data, the ratios are defined with respect to the data.

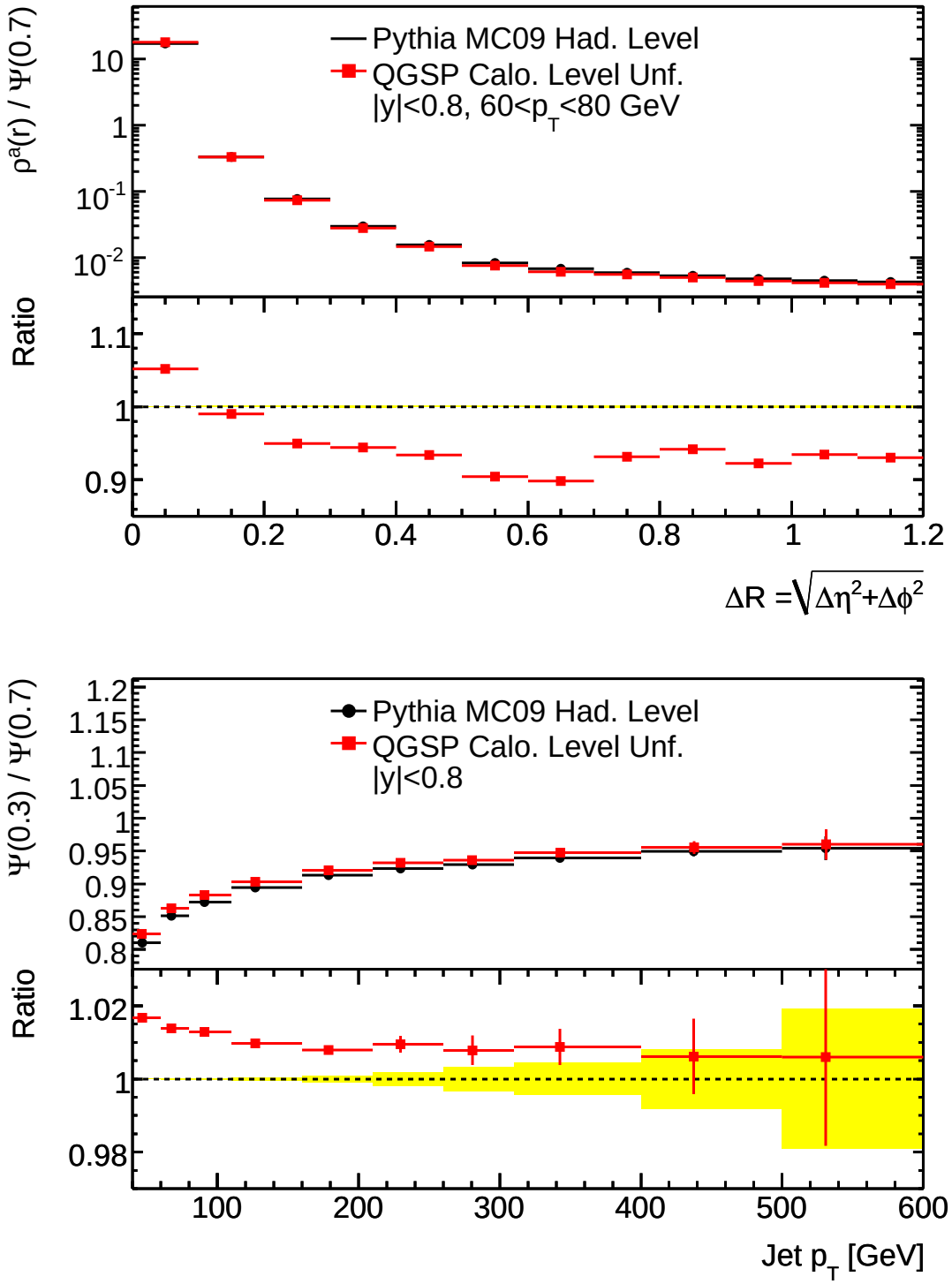


Figure A.1: Comparison of the jet shapes $\rho^a(r)/\Psi(0.7)$ (top) and $\Psi(0.3)/\Psi(0.7)$ (bottom) with the QGSP physics list after unfolding using the QGSP_BERT physics list.

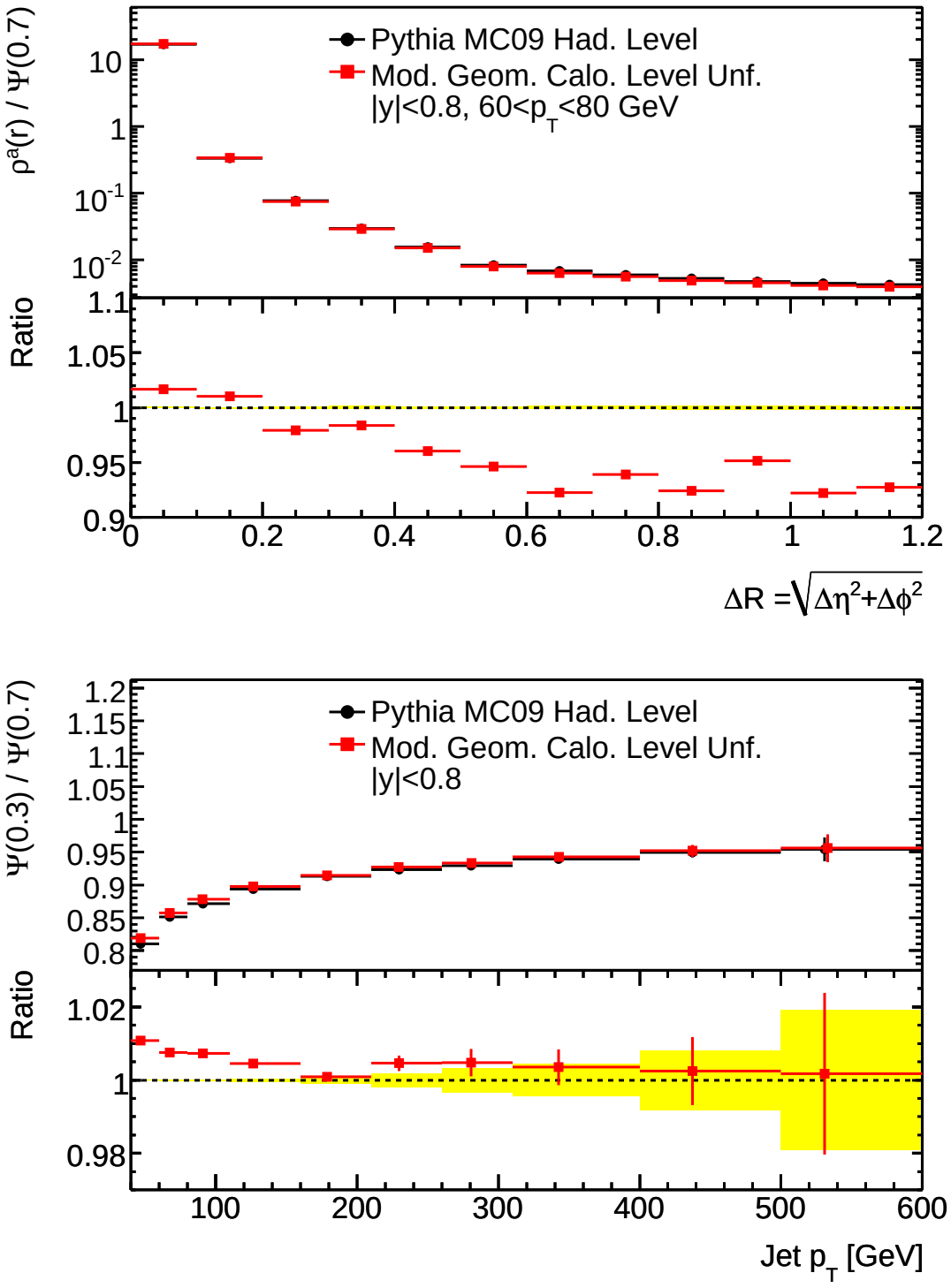


Figure A.2: Comparison of jet shapes with and without additional material in the LAr calorimeter.

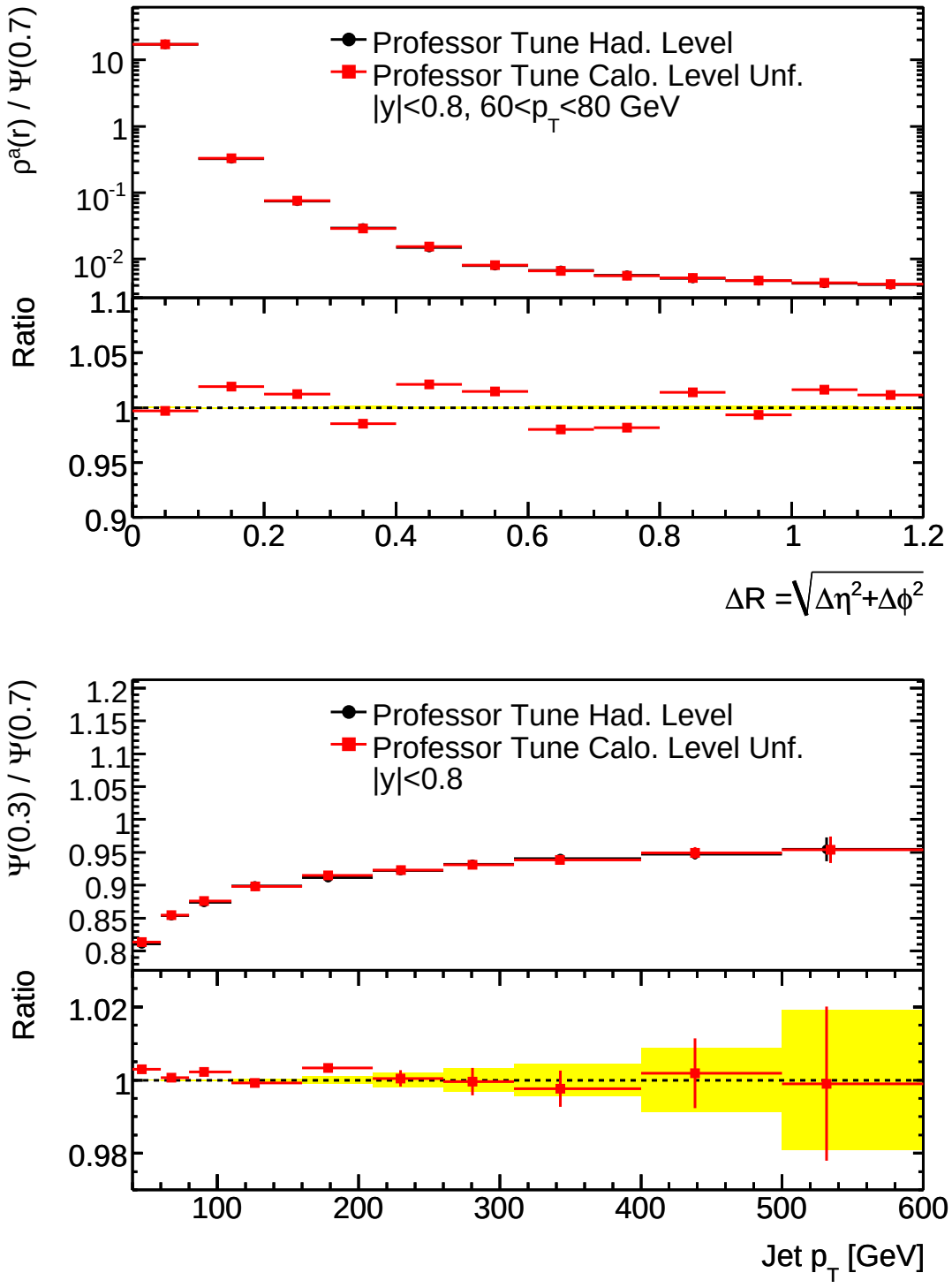


Figure A.3: Comparison of jet shapes in Monte Carlo generated with the MC09 and Professor tunes.

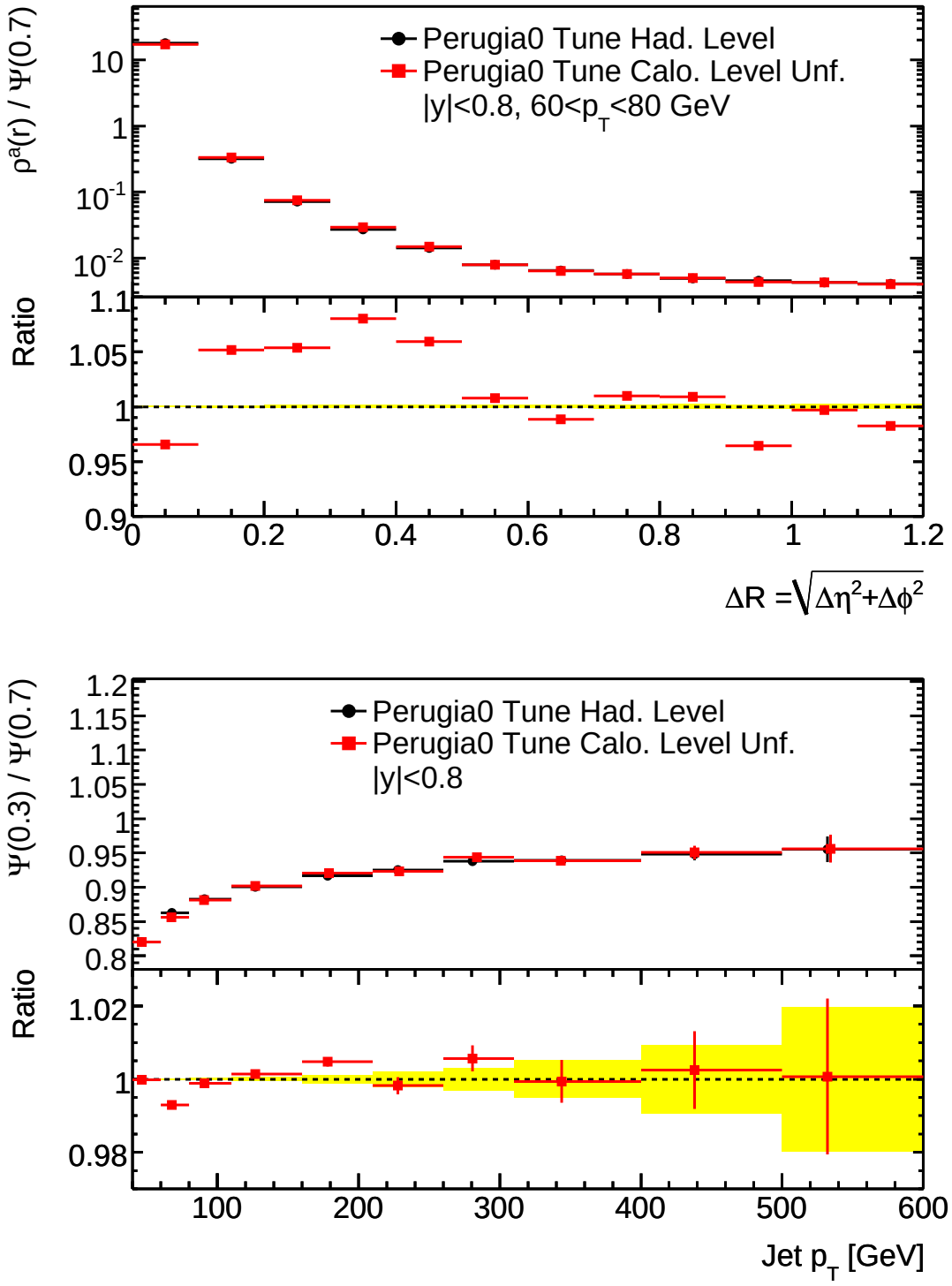


Figure A.4: Comparison of jet shapes in Monte Carlo generated with a different underlying event tune and with a different parton distribution function.

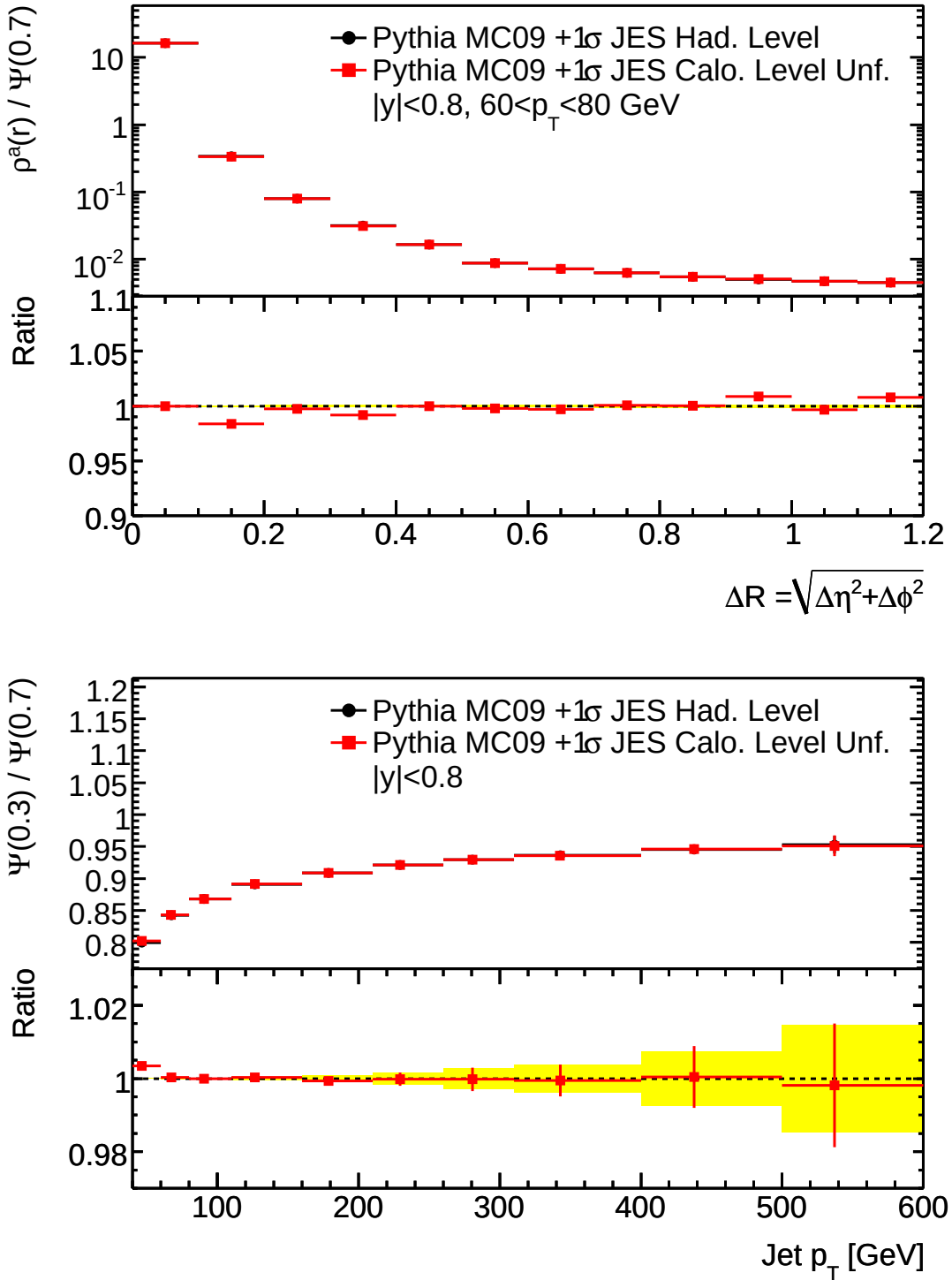


Figure A.5: Jet shapes for samples with modified kinematics after unfolding. The bias caused by a harder p_T spectrum in data is approximated by the difference between the black and red points in the two plots.

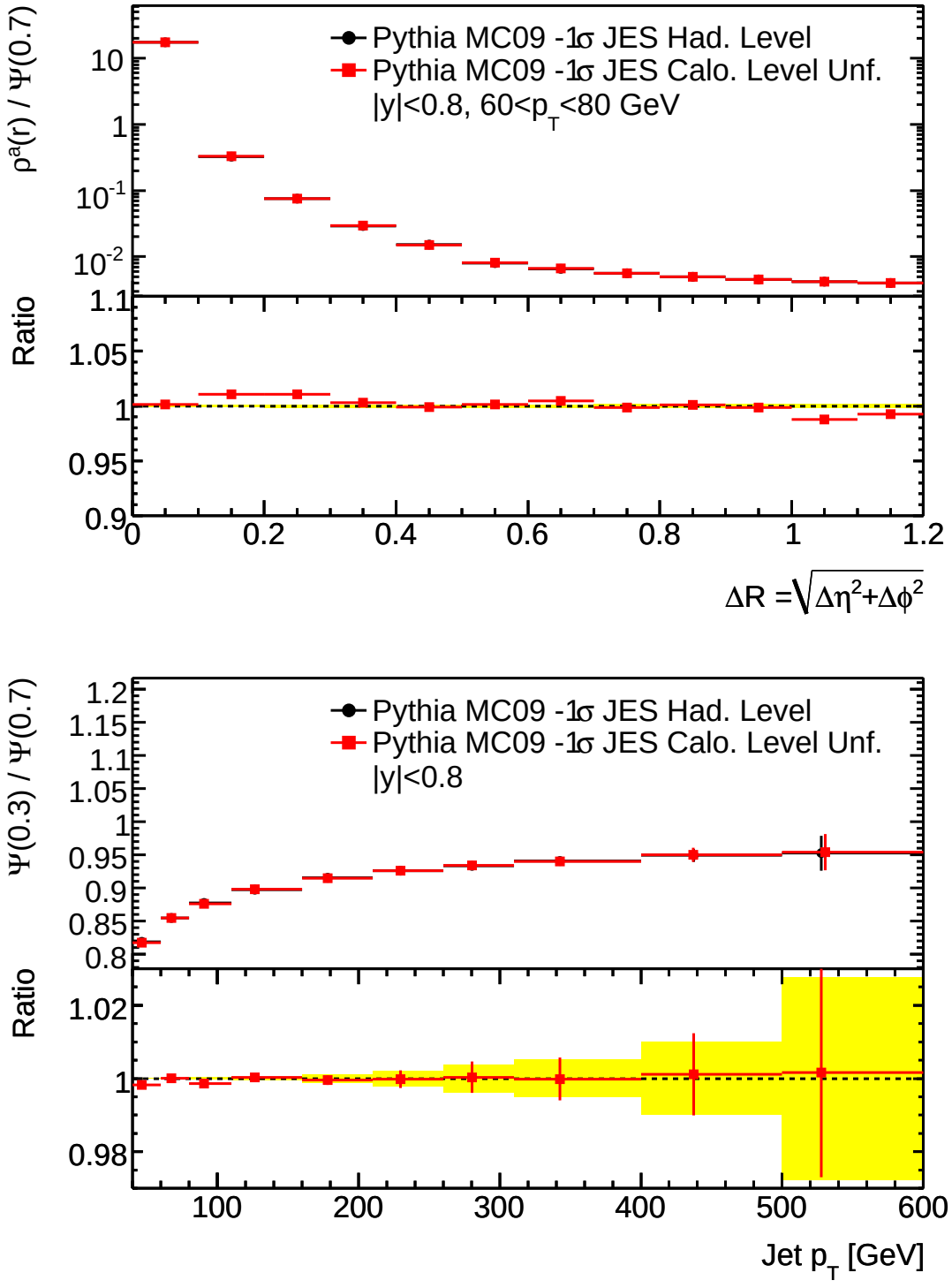


Figure A.6: Jet shapes for samples with modified kinematics after unfolding. The bias caused by a softer p_T spectrum in data is approximated by the difference between the black and red points in the two plots.

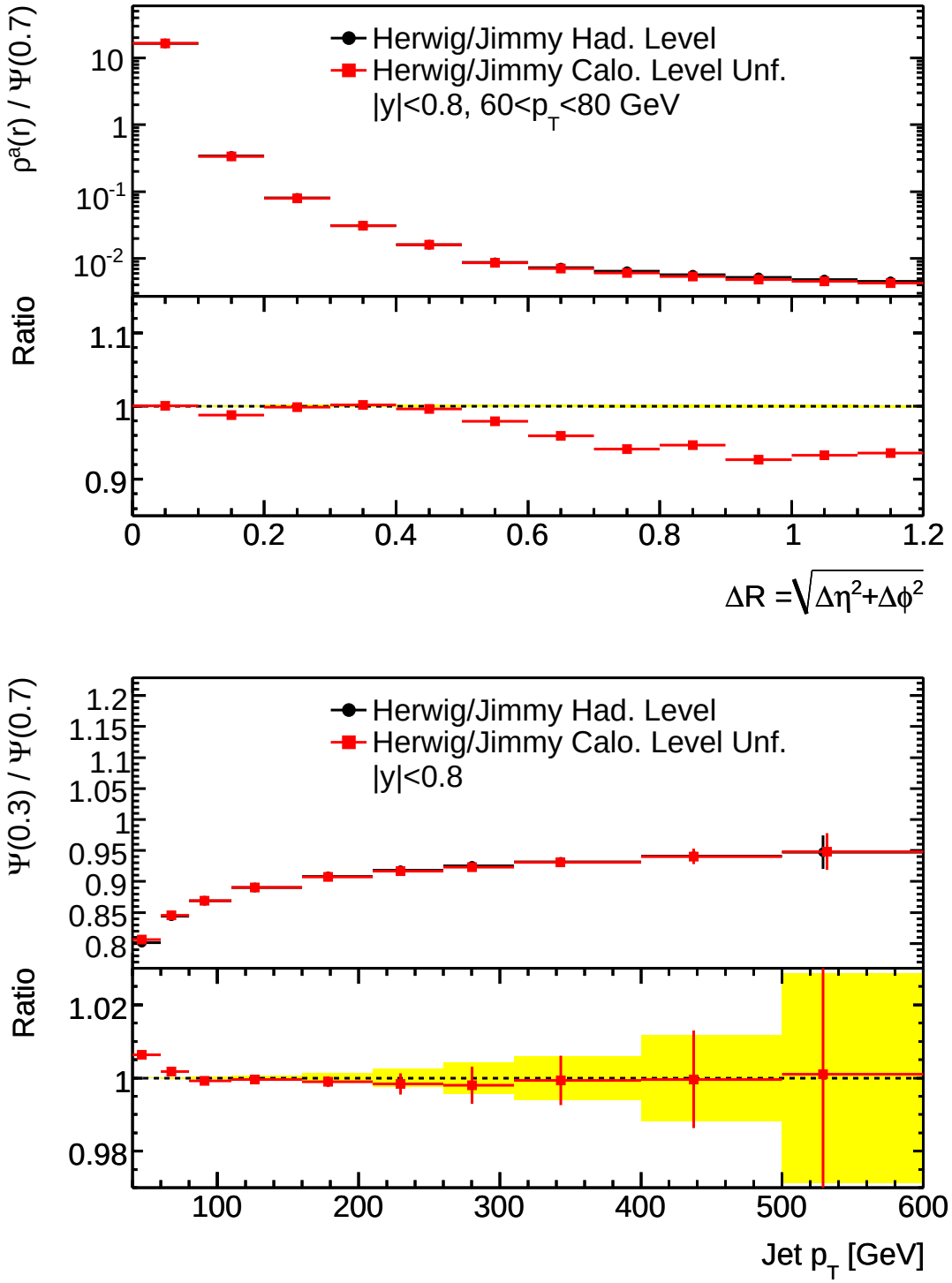


Figure A.7: The jet shapes $\rho^a(r)/\Psi(0.7)$ (top) and $\Psi(0.3)/\Psi(0.7)$ (bottom) in Herwig/Jimmy at the hadron level and after unfolding based on Pythia's MC09 tune.

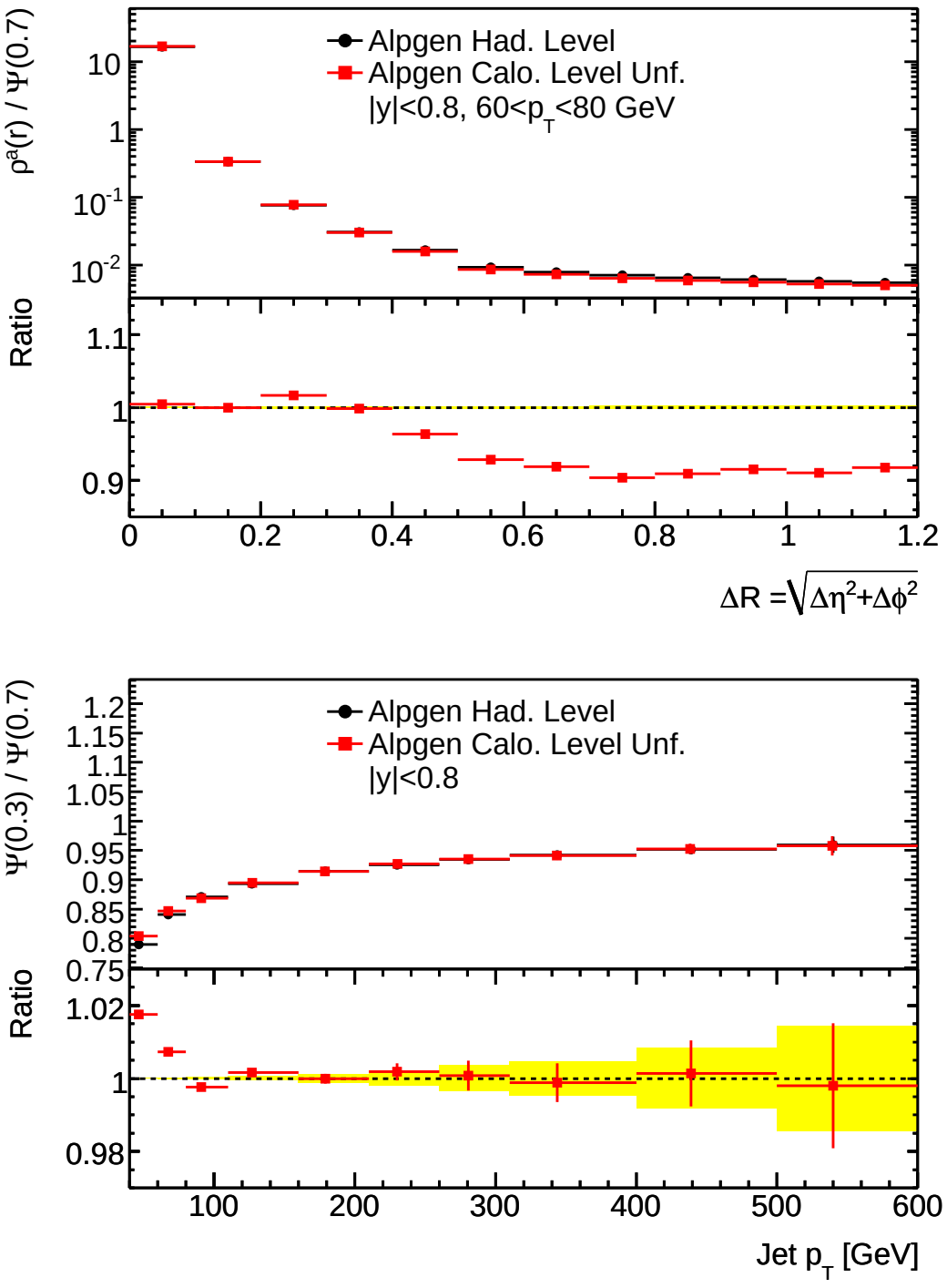


Figure A.8: Jet shapes in ALPGEN at the hadron level and after unfolding. Whereas PYTHIA and HERWIG model only 2→2 processes at the matrix-element level, ALPGEN includes processes up to 2→6. The hadronization model of HERWIG and the underlying event model of JIMMY are then used to fill in the remainder of the event.

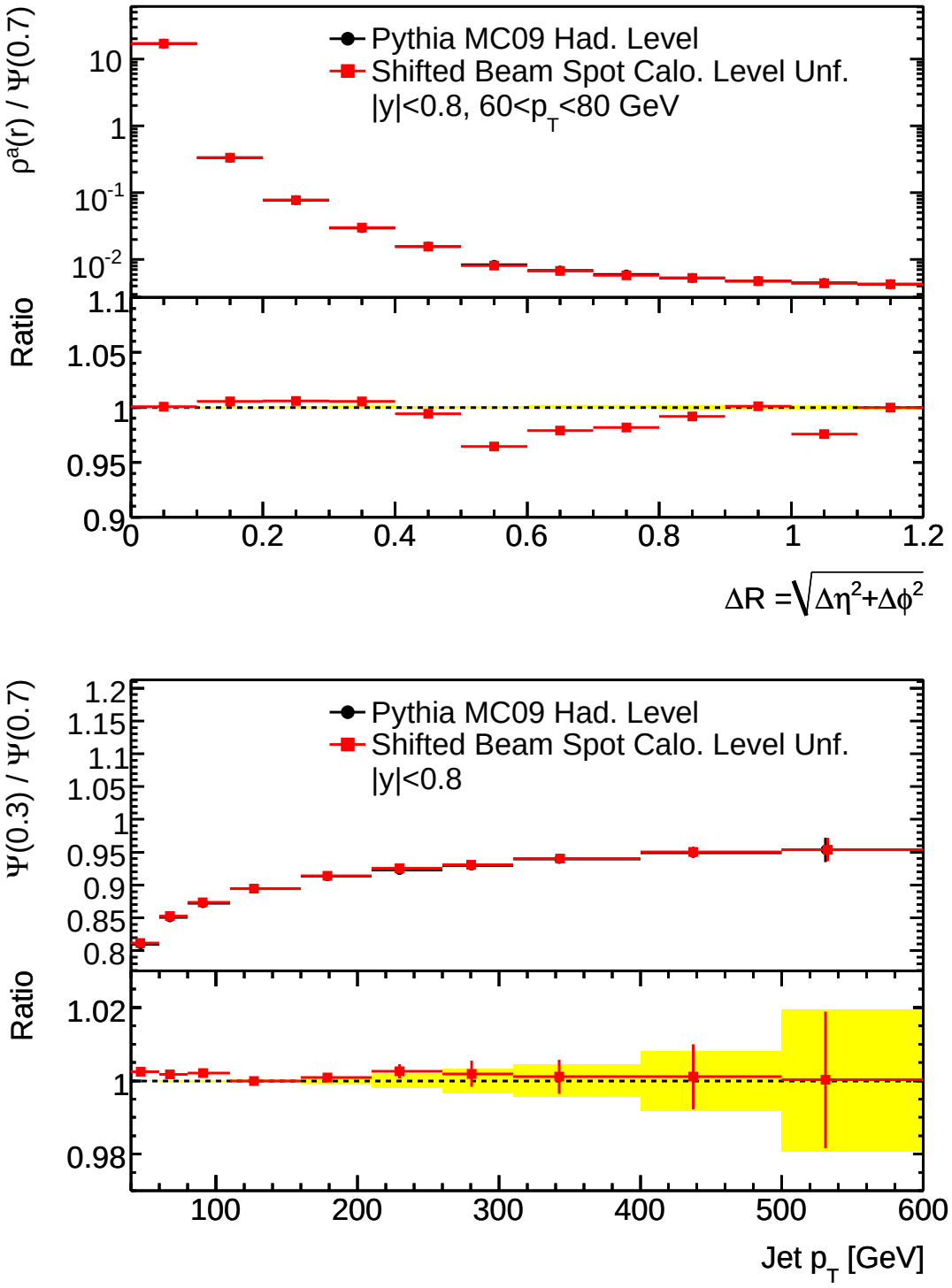


Figure A.9: Comparison of unfolded reconstructed and hadron-level jet shapes in events with a displaced luminous region.

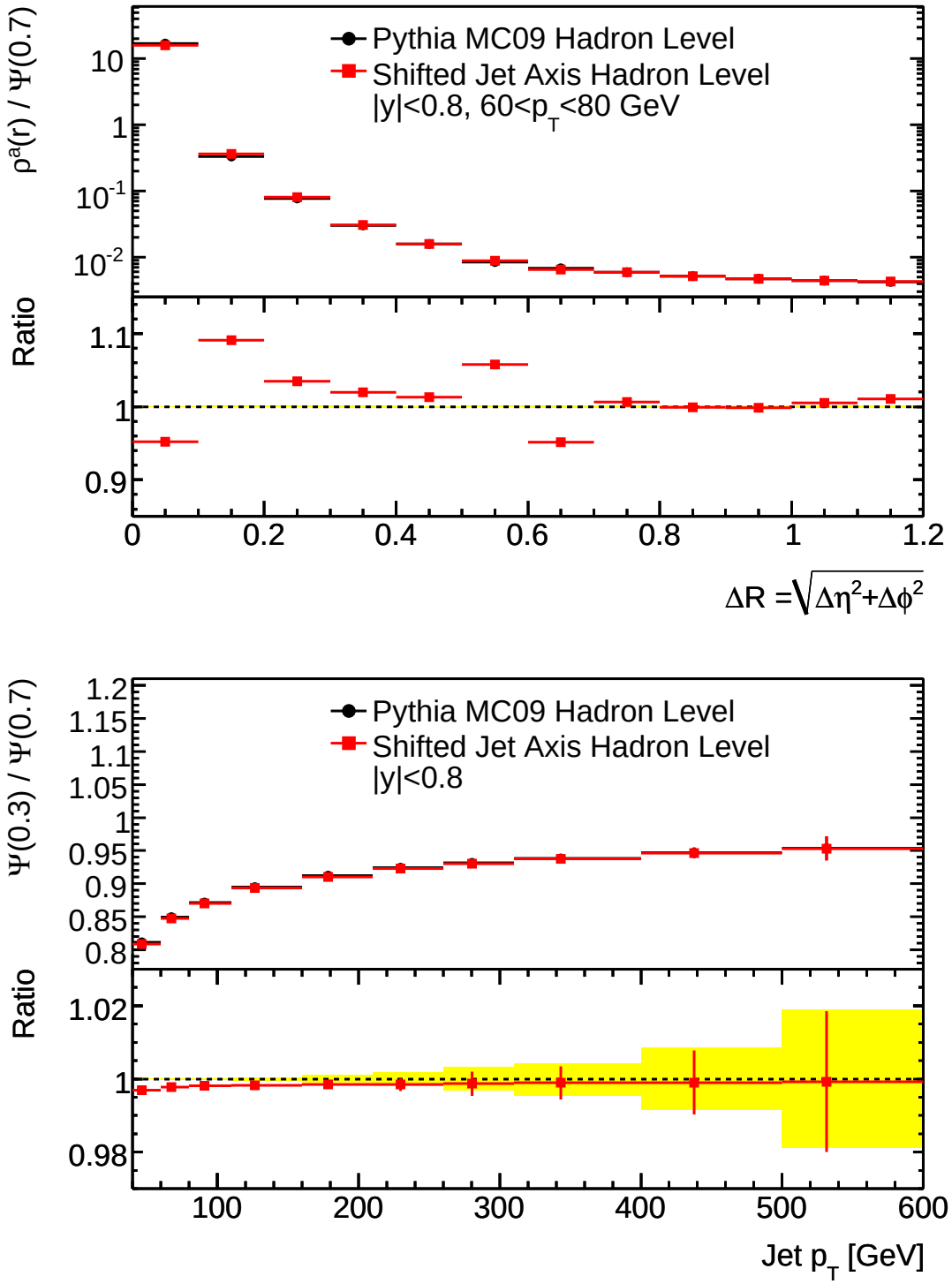


Figure A.10: Comparison of hadron-level jet shapes in jets with and without a smearing of the jet axis.

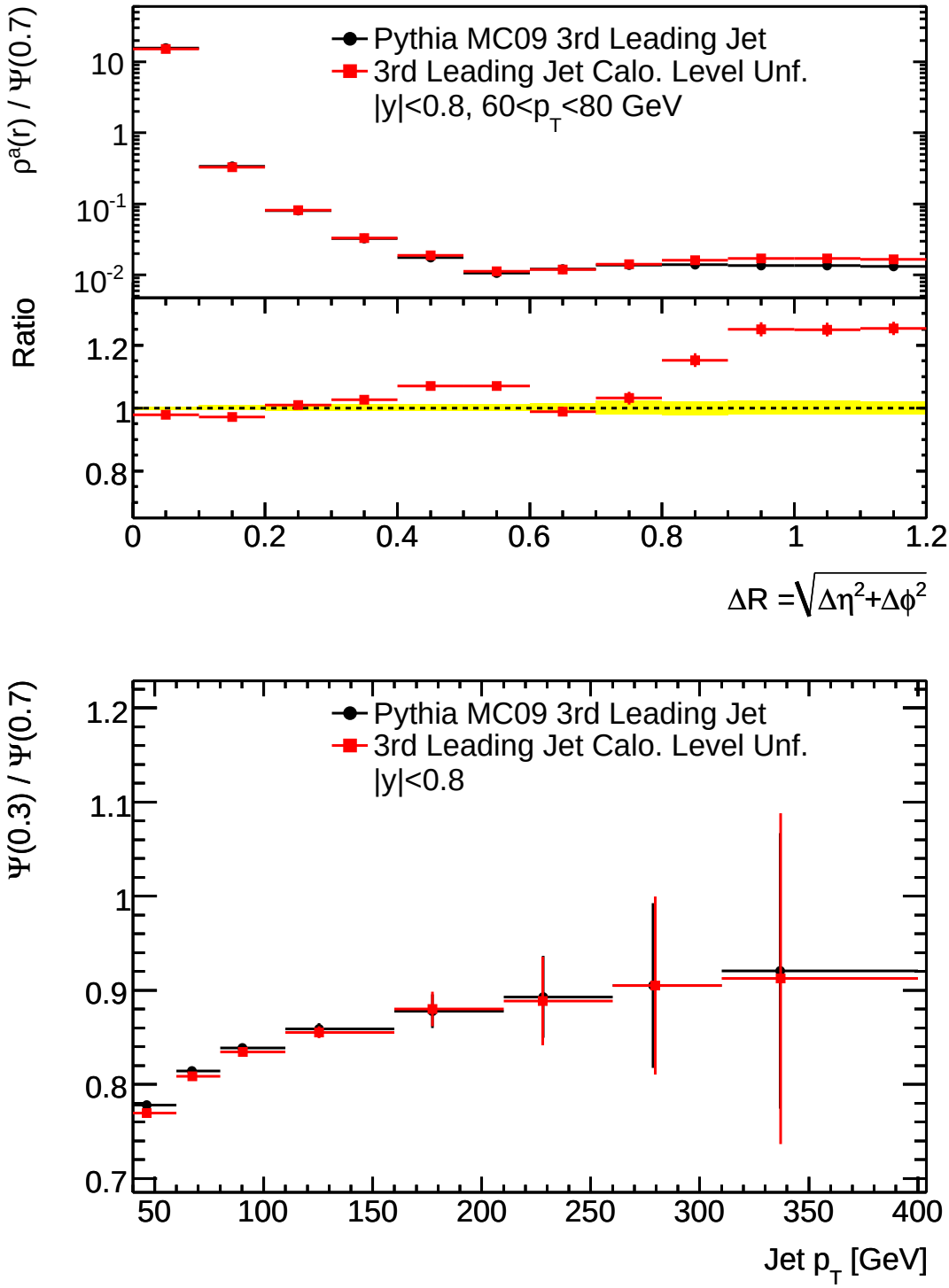


Figure A.11: Comparison of 3rd leading jet shapes in Monte Carlo at the hadron level and after detector simulation and unfolding. The unfolding is based on the inclusive sample.

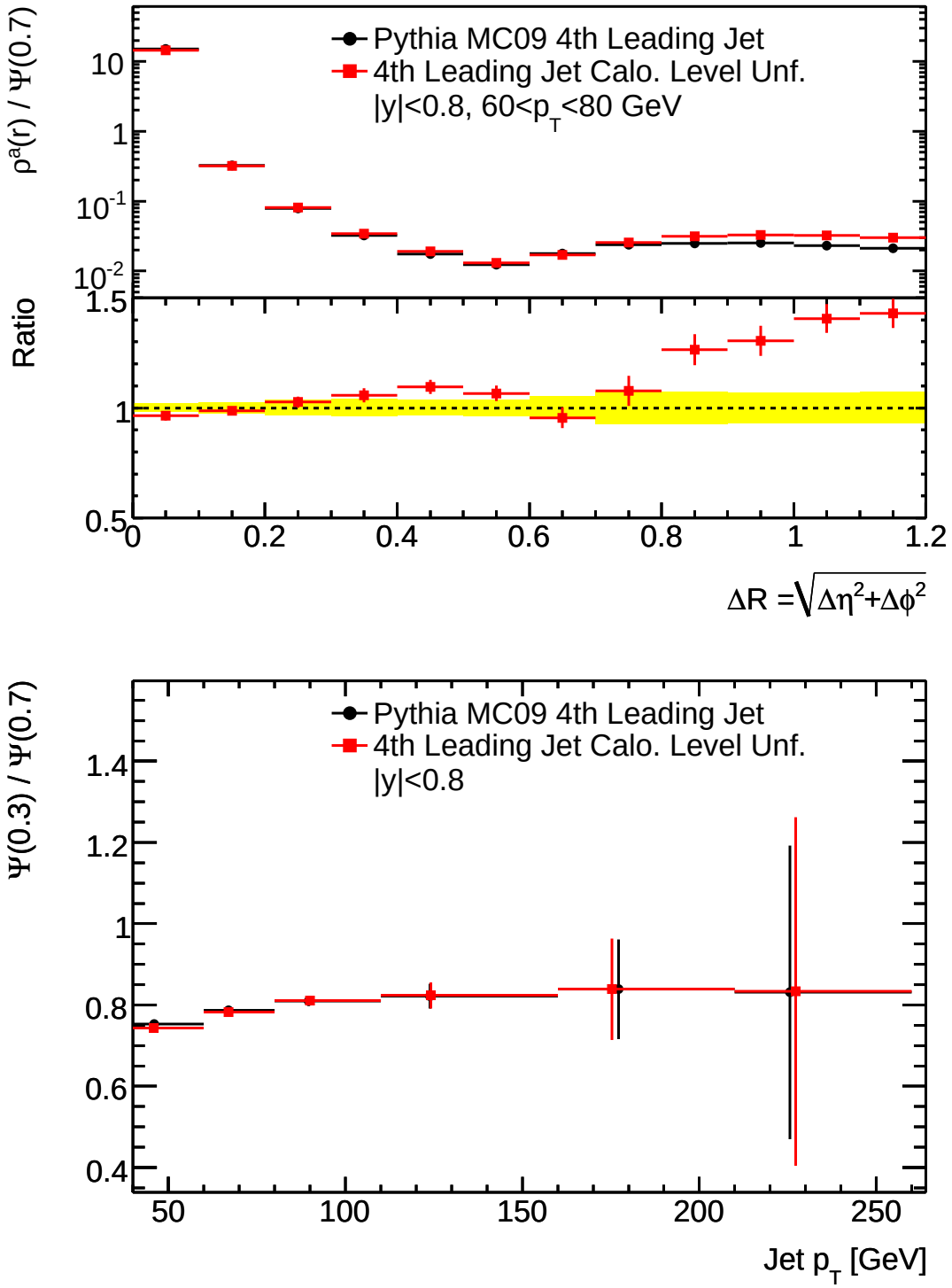


Figure A.12: Comparison of 4th leading jet shapes in Monte Carlo at the hadron level and after detector simulation and unfolding. The unfolding is based on the inclusive sample.

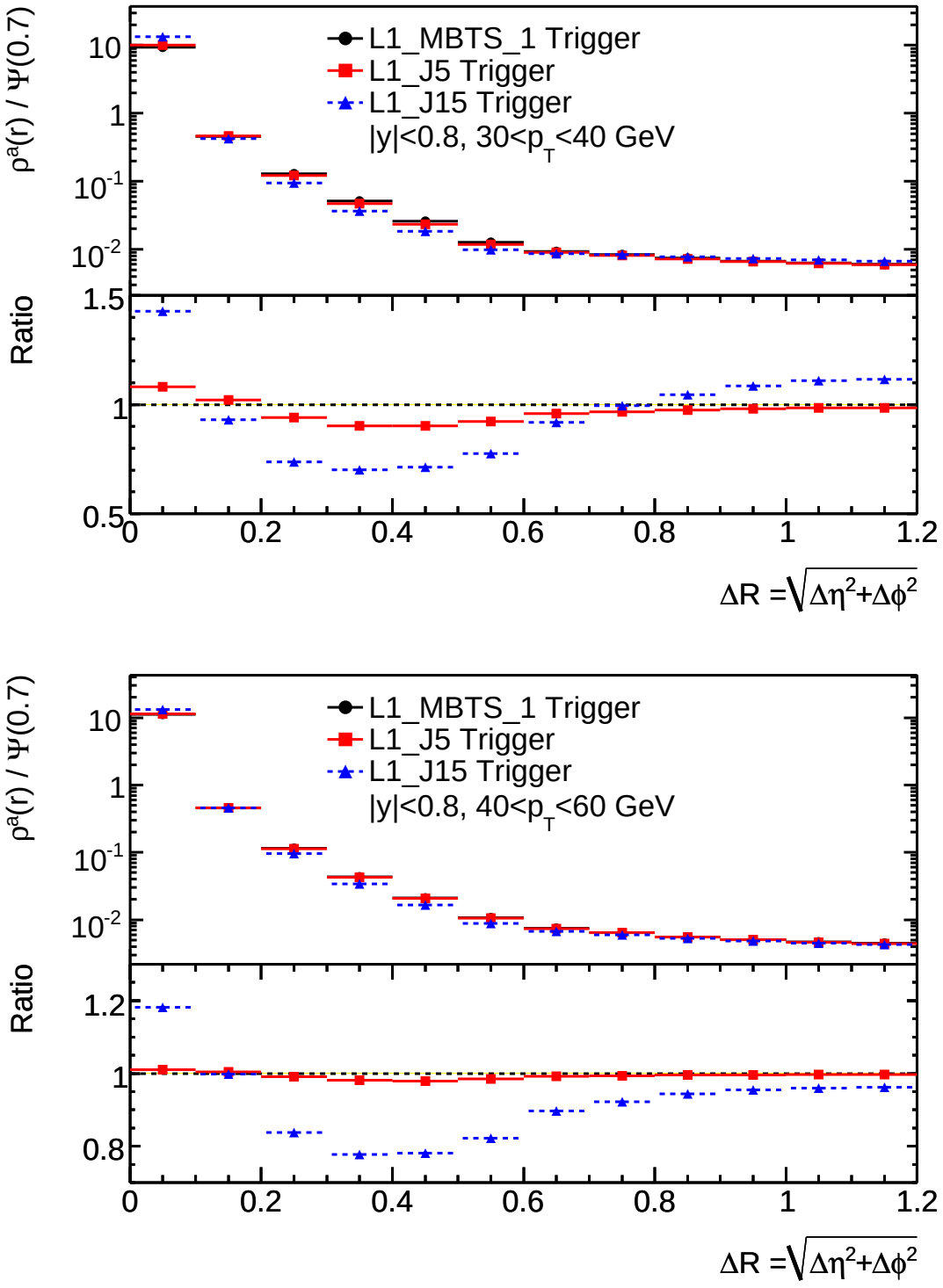


Figure A.13: The jet shape variable $\rho^a(r)/\Psi(0.7)$ for two p_T bins near the threshold of the L1 jet triggers in Monte Carlo simulation. The bias is dramatically reduced at higher p_T.

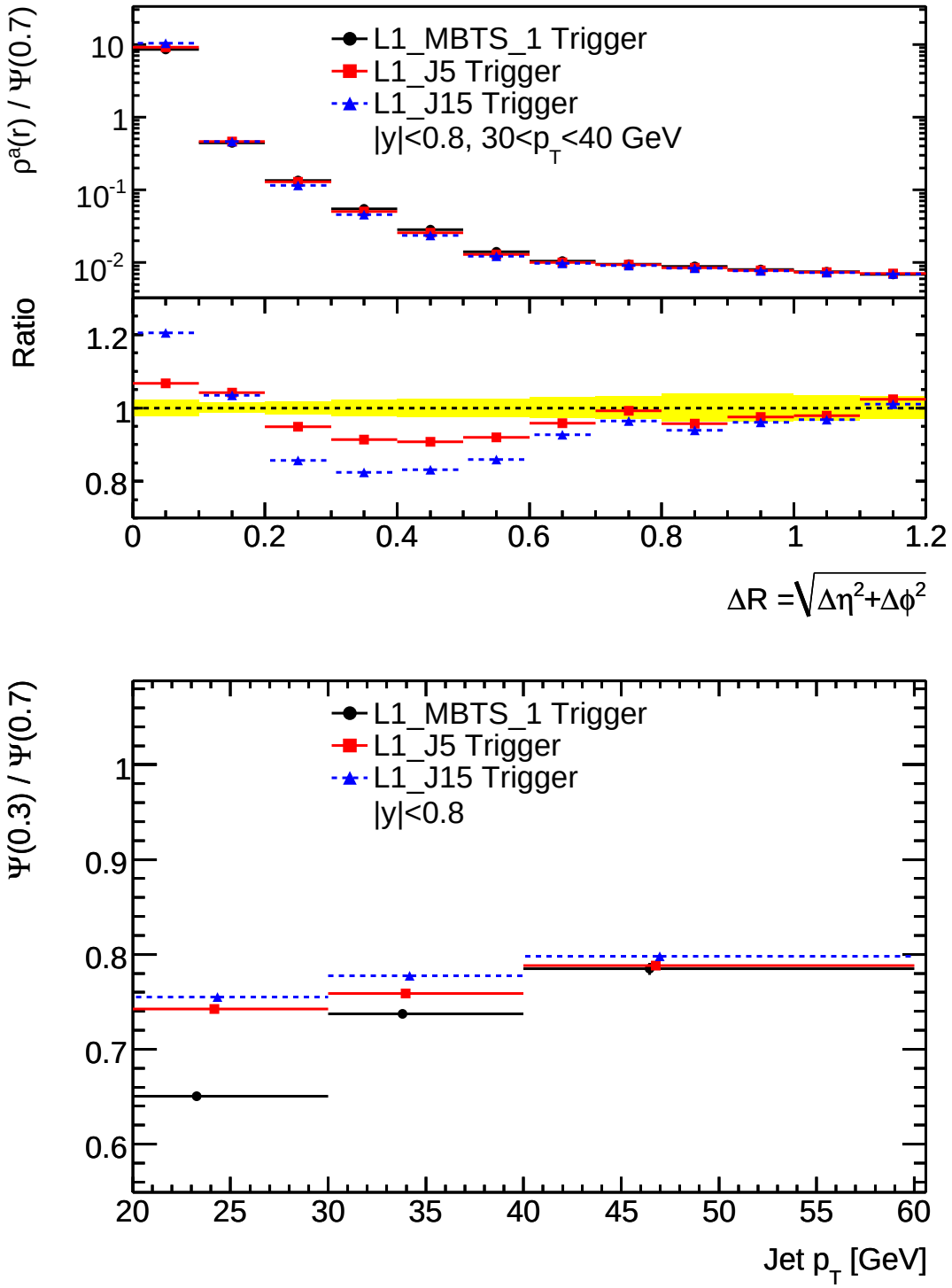


Figure A.14: The jet shapes in data using several different trigger configurations in data. The calorimeter triggers have a bias at low p_T.

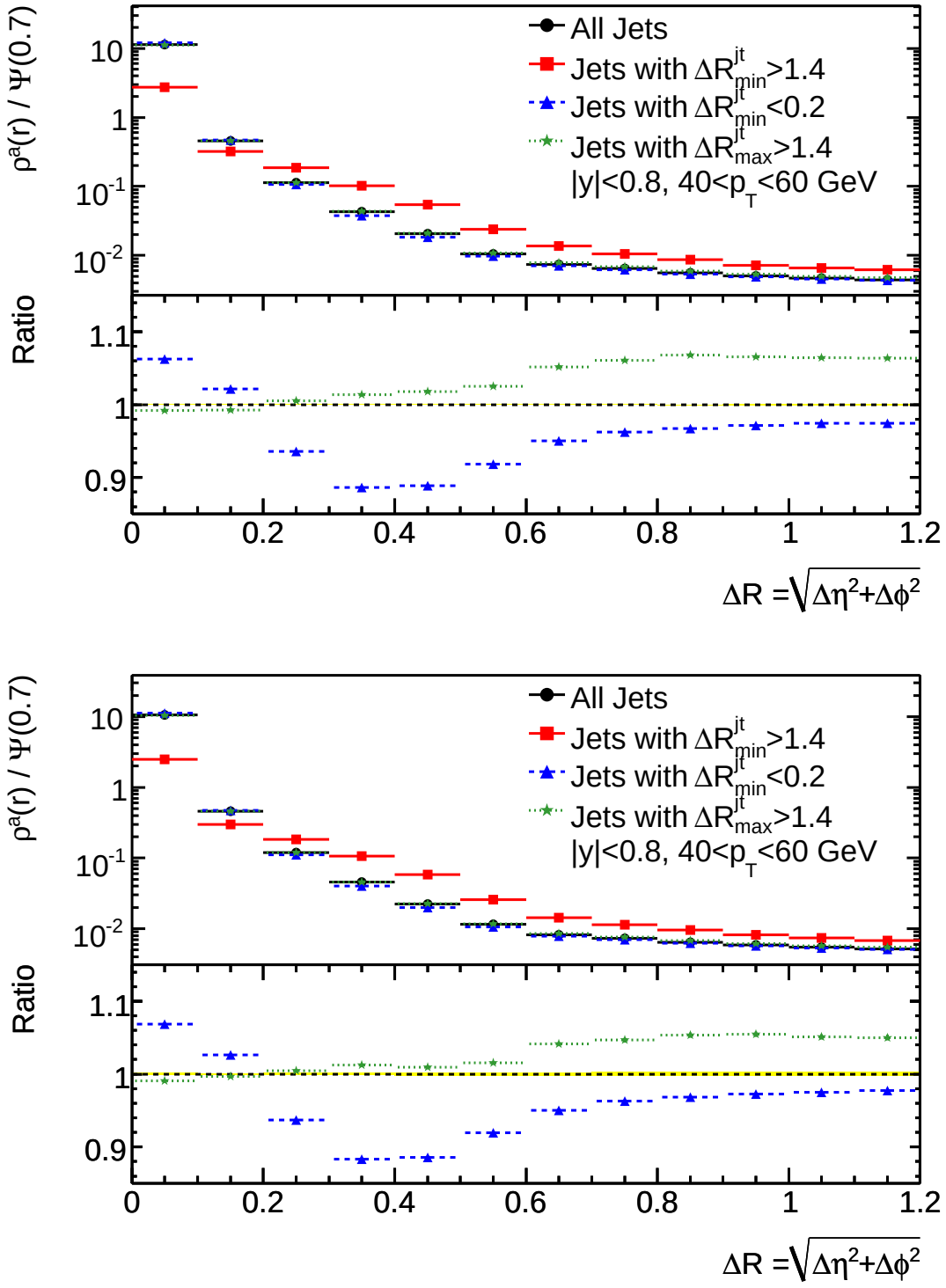


Figure A.15: Comparison of jet shapes in jets which overlapped in various ways with a trigger jet in Monte Carlo (top) and data (bottom).

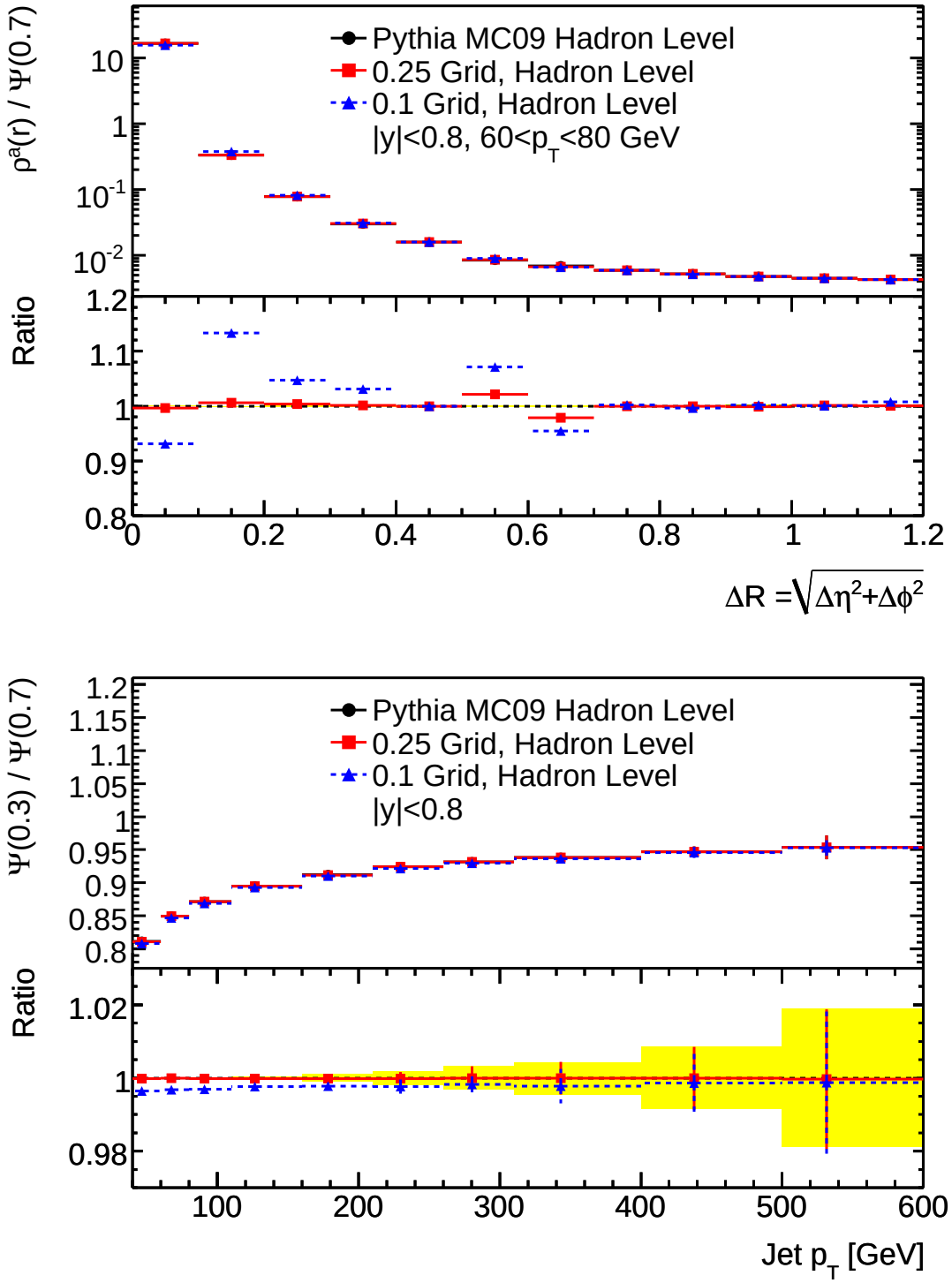


Figure A.16: Comparison of jet shapes constructed using stable hadrons with and without collection into a grid with granularity of 0.025 units (red) and 0.1 units (blue) in y and ϕ.

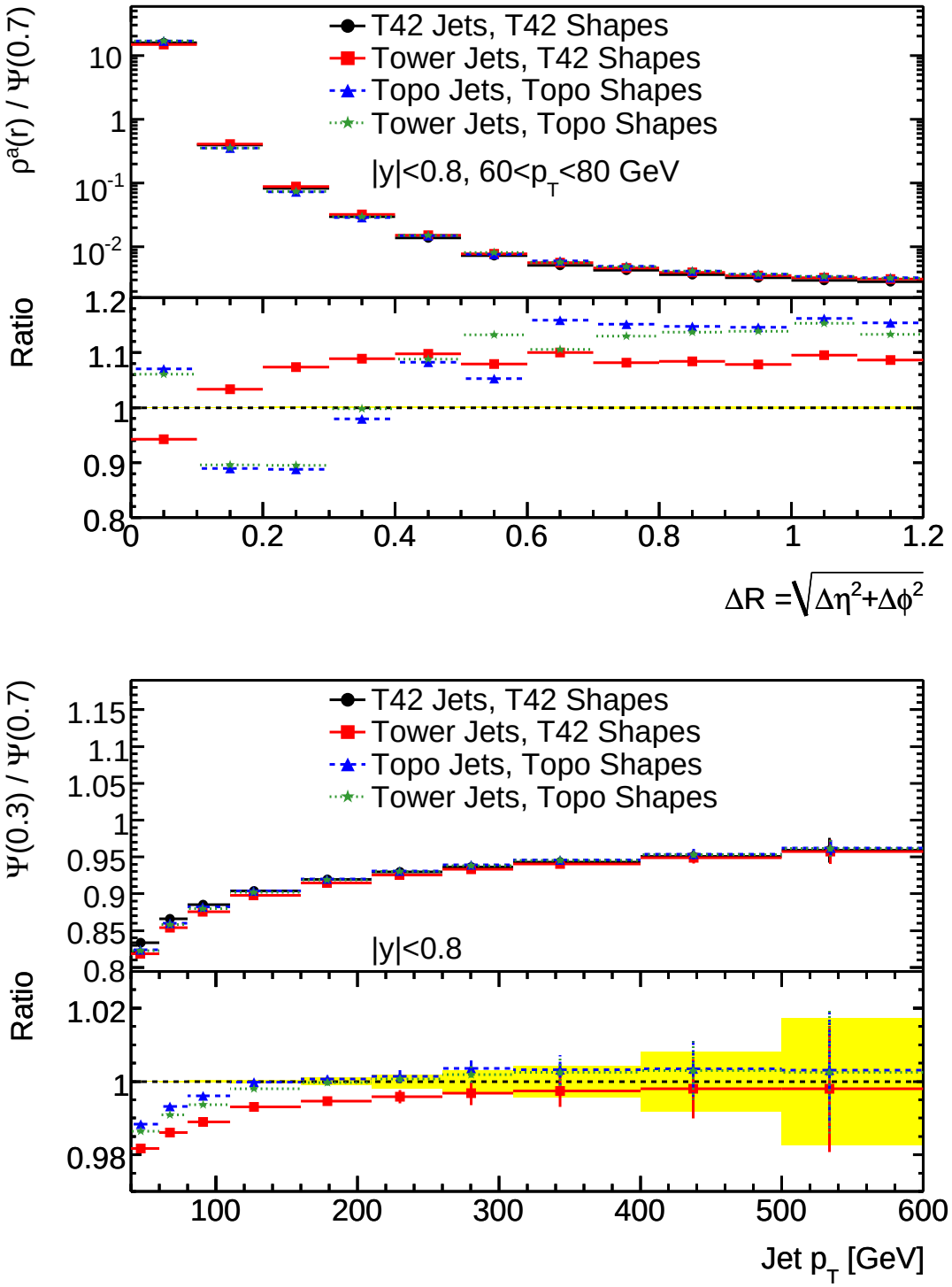


Figure A.17: Comparison of jet shapes for towers with standard and "T42" topological noise suppression.

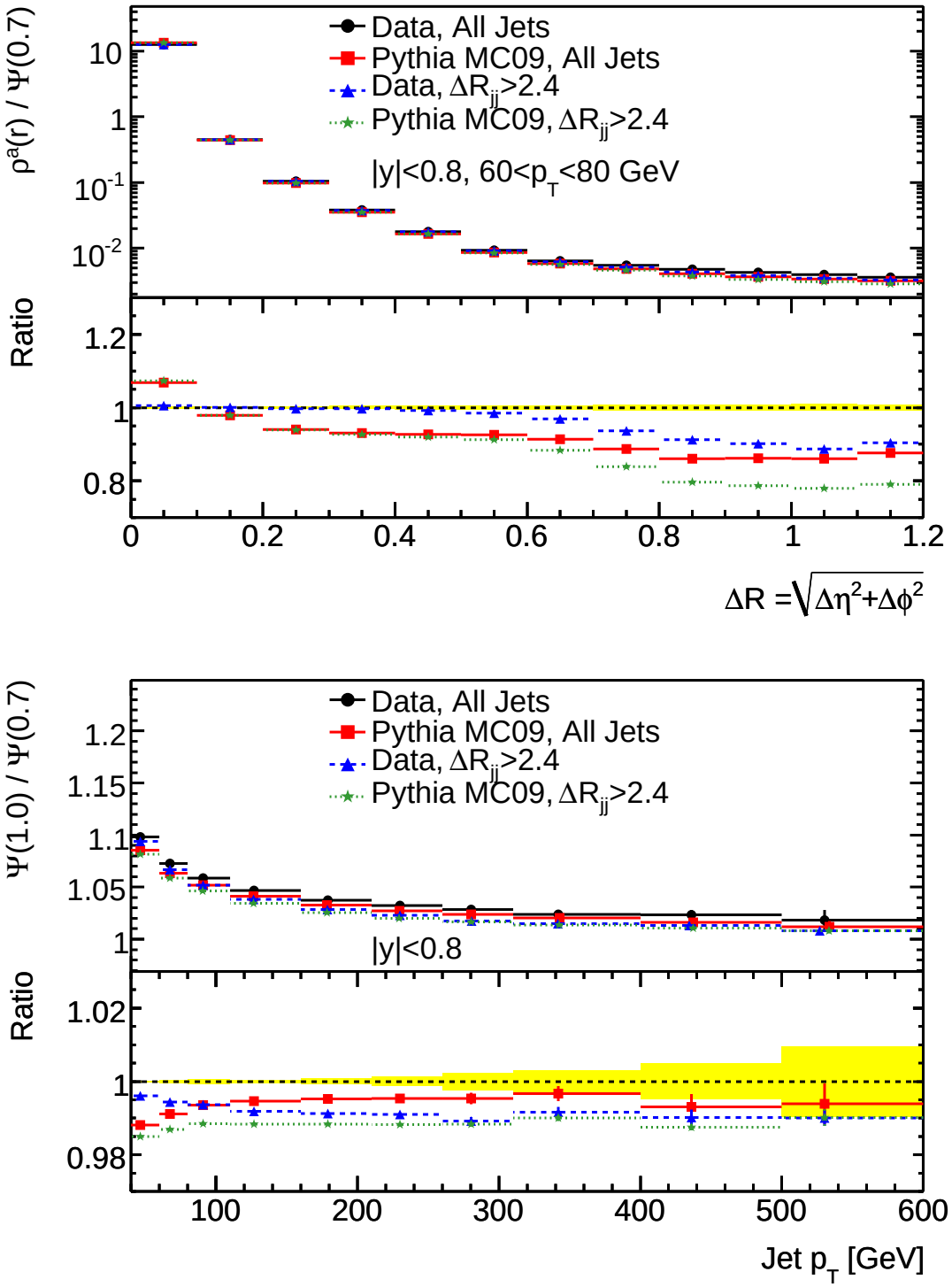


Figure A.18: Comparison of jet shapes in jets which were far from another jet (ΔR>2.4) and all jets.

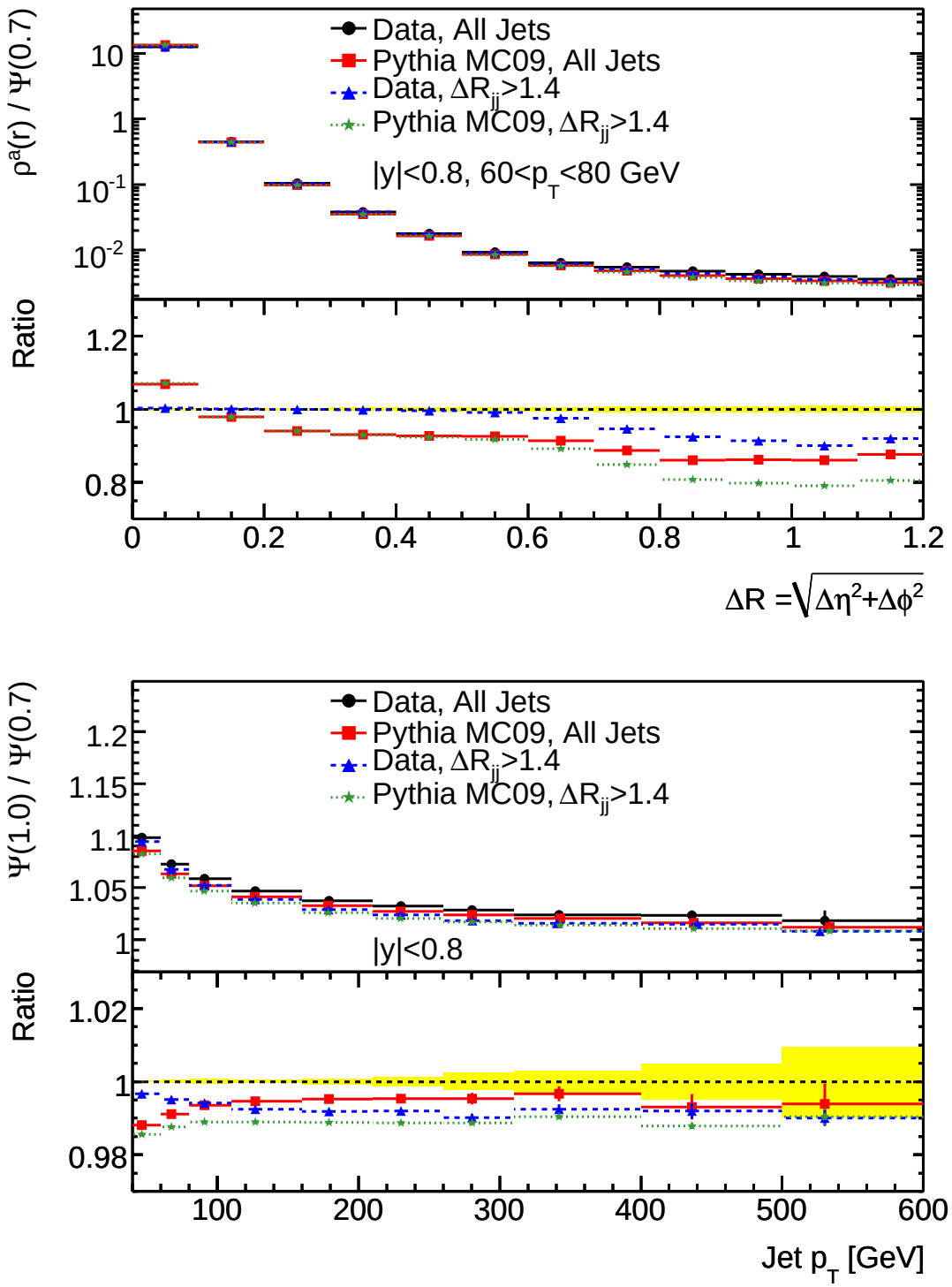


Figure A.19: Comparison of jet shapes in jets which did not overlap with another jet (ΔR_{jj} >1.4) and all jets.

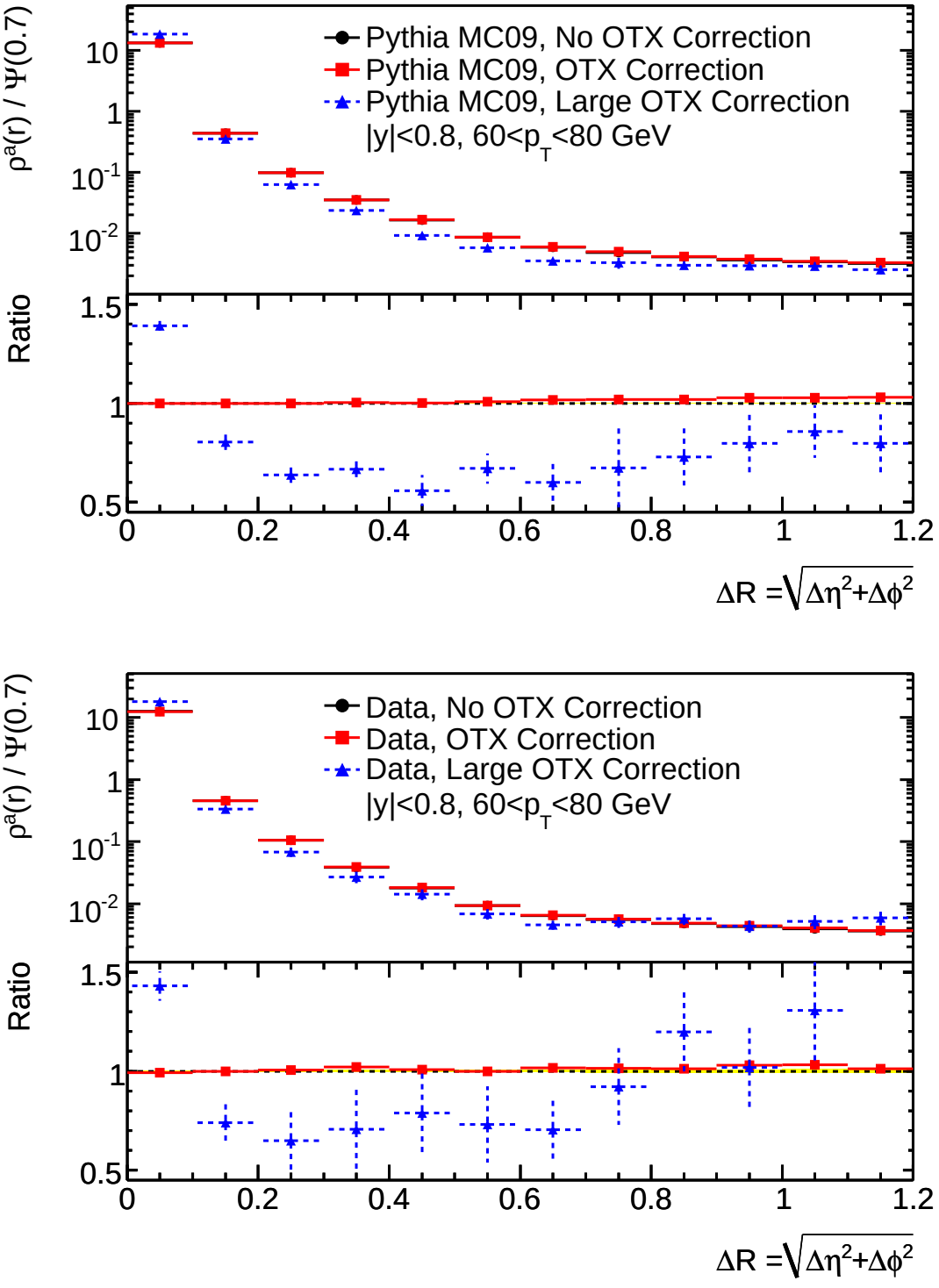


Figure A.20: Comparison of jet shapes for jets with no OTX correction, all jets with the correction, and only jets with a large OTX correction (>20%). Jets with a large correction are significantly wider.

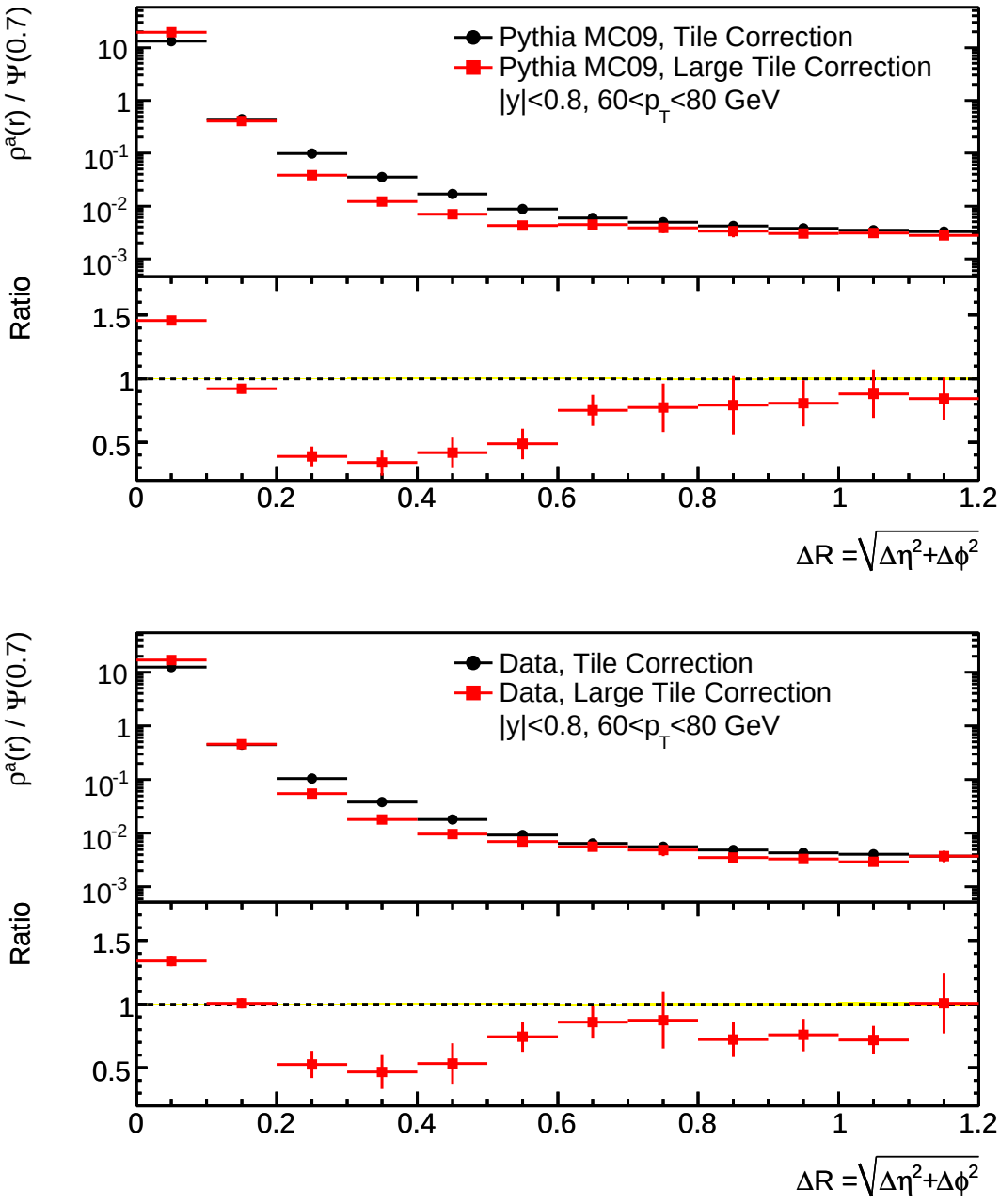


Figure A.21: Comparison of jet shapes for all jets and jets with large tile calorimeter ("nearest neighbor") correction (>20%). Jets with a large correction are significantly wider.

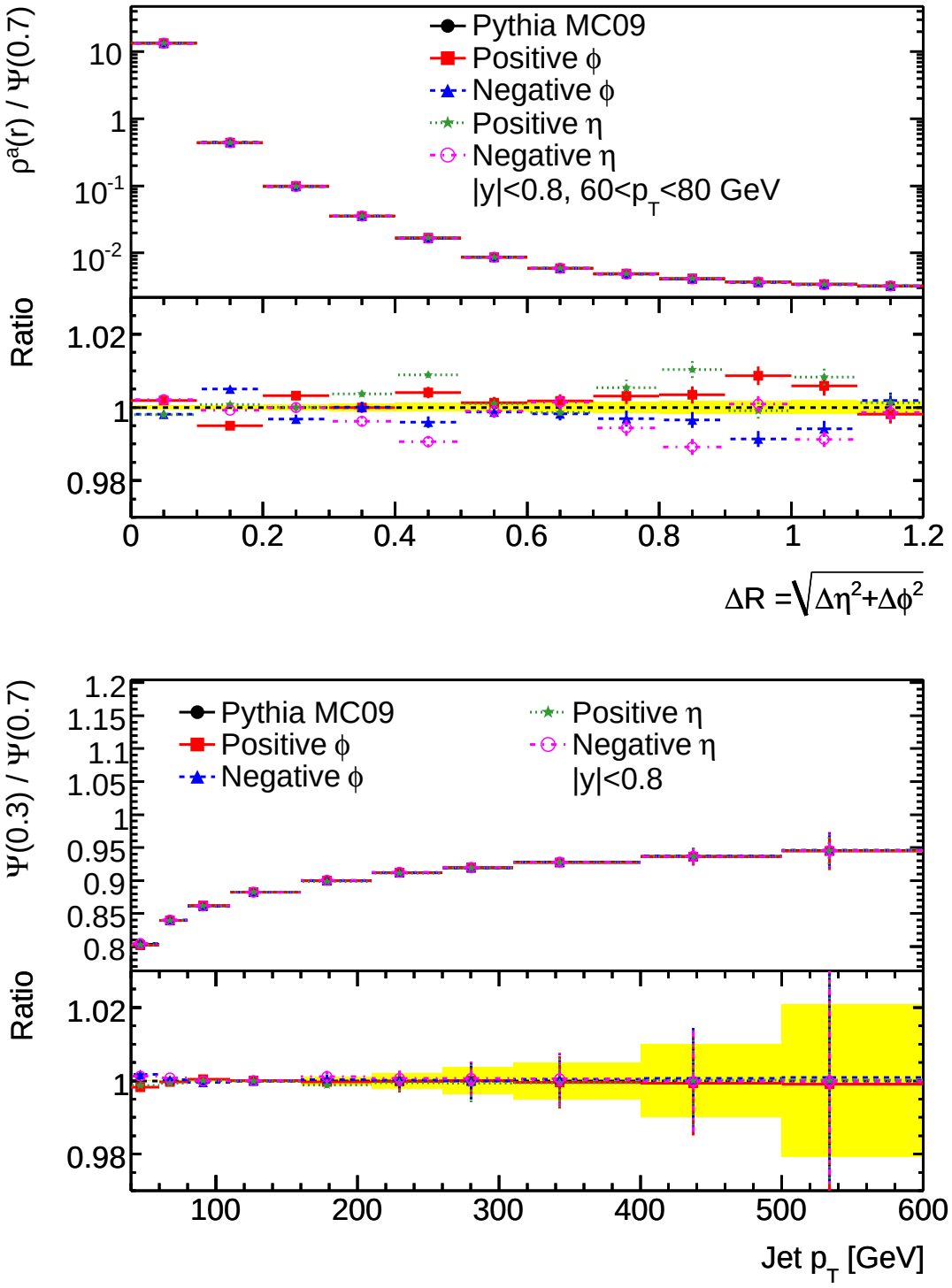


Figure A.22: Comparison of jet shapes for all jets, jets with positive and negative ϕ, and jets with positive and negative y, in Monte Carlo simulation.

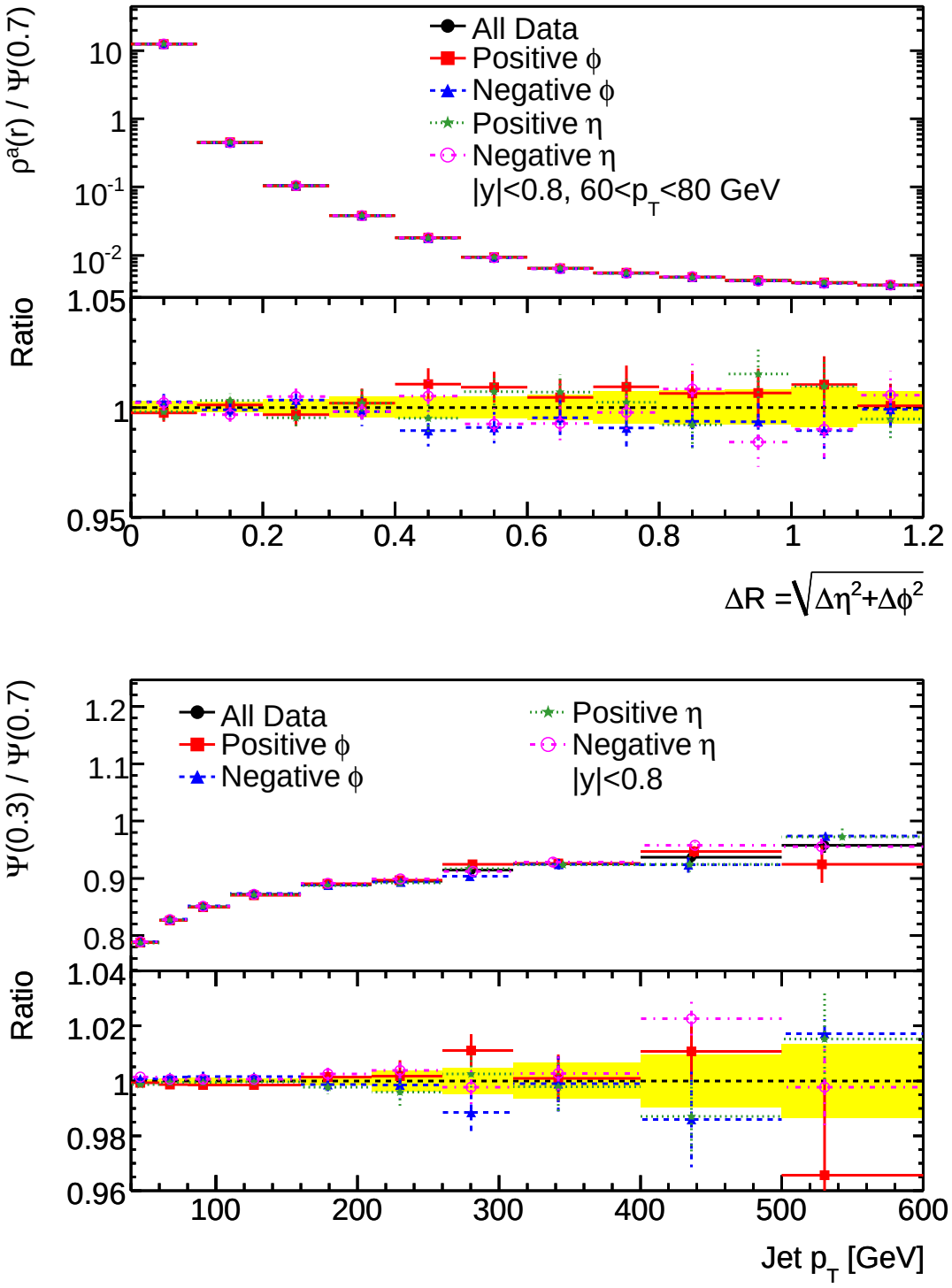


Figure A.23: Comparison of jet shapes for all jets, jets with positive and negative ϕ, and jets with positive and negative y, in the data.

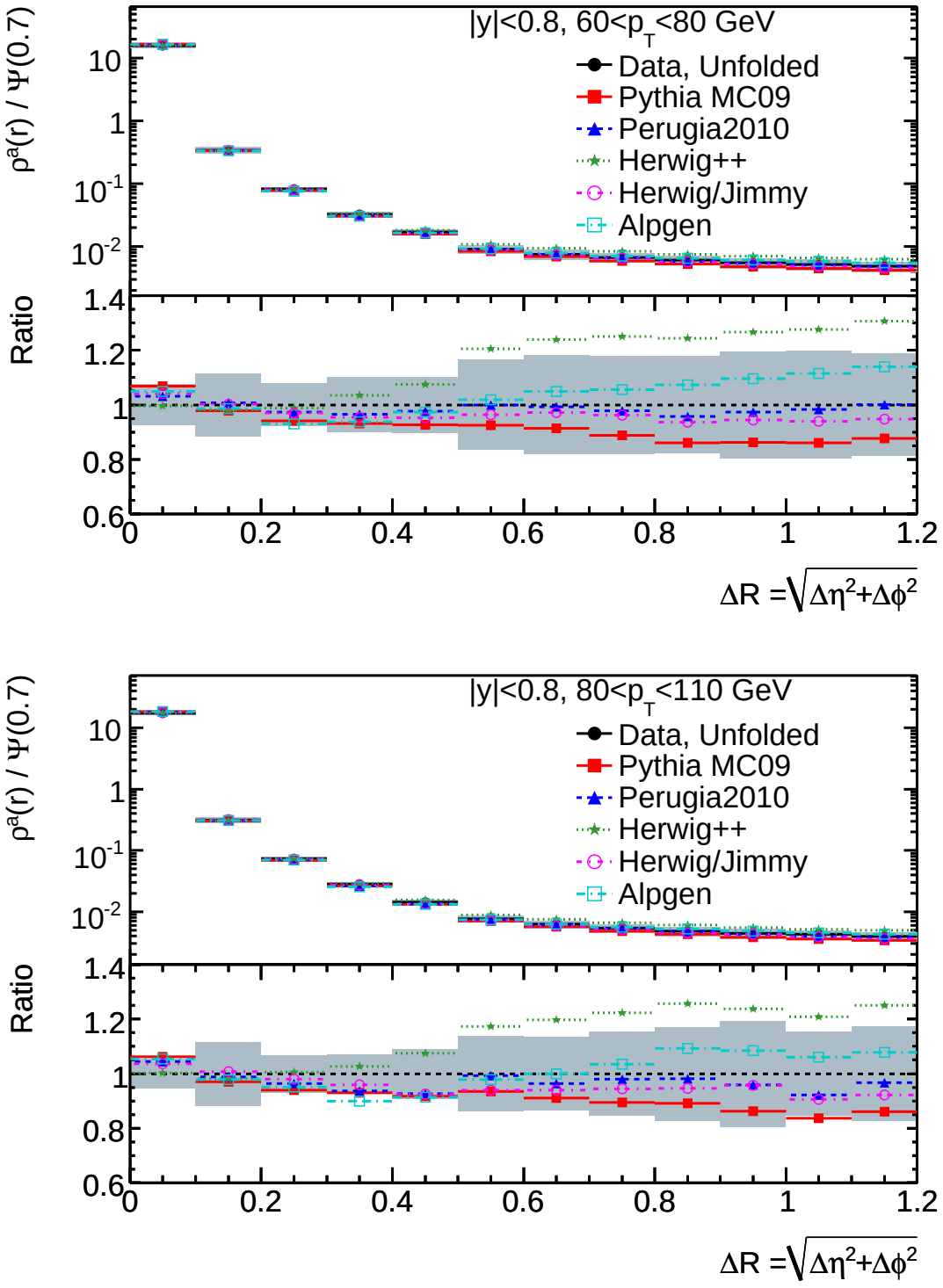


Figure A.24: Differential jet shapes after unfolding, for jets with $|y| < 0.8$, in two p_T bins.

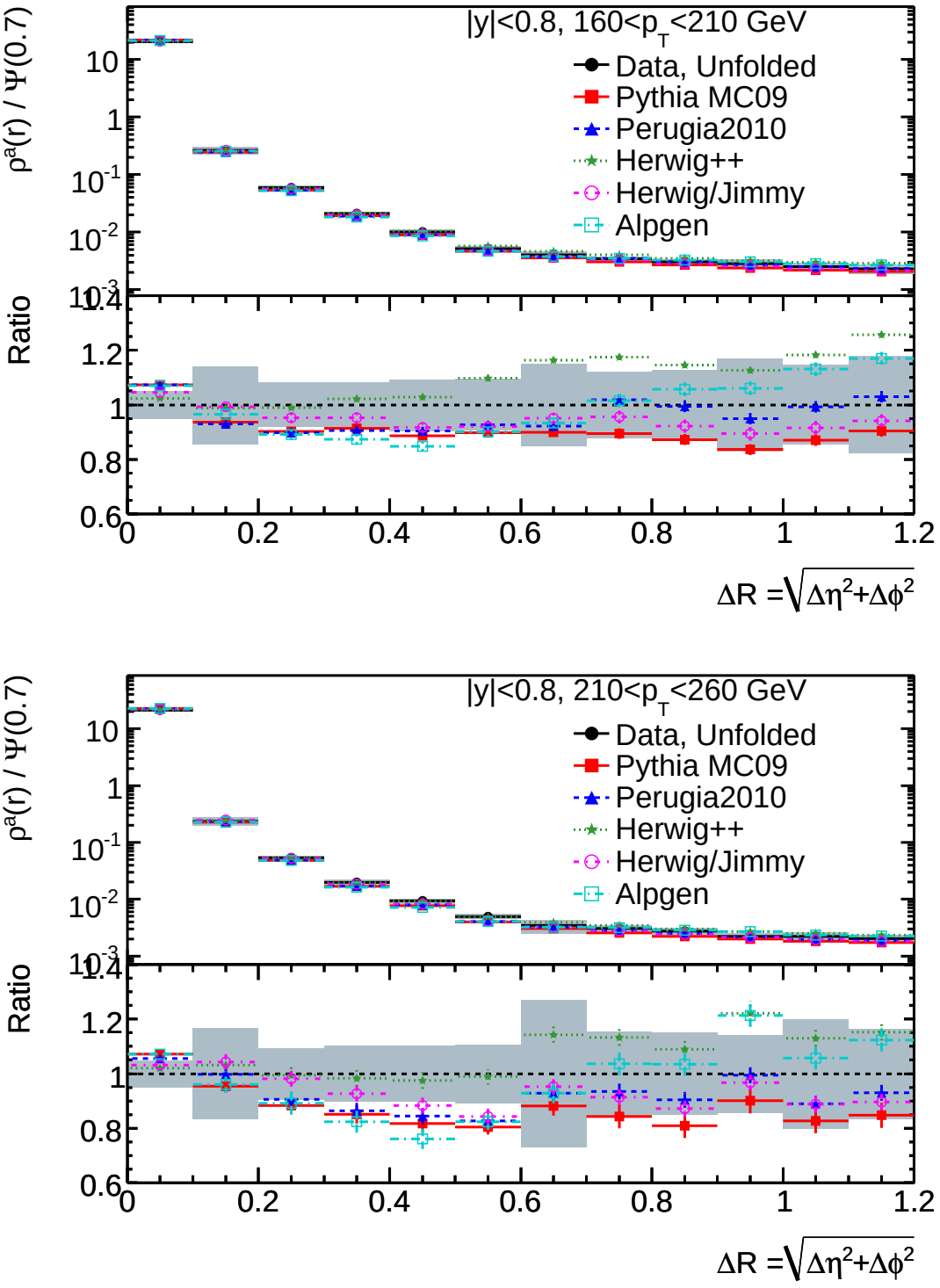


Figure A.25: Differential jet shapes after unfolding, for jets with $|y| < 0.8$, in two p_T bins.

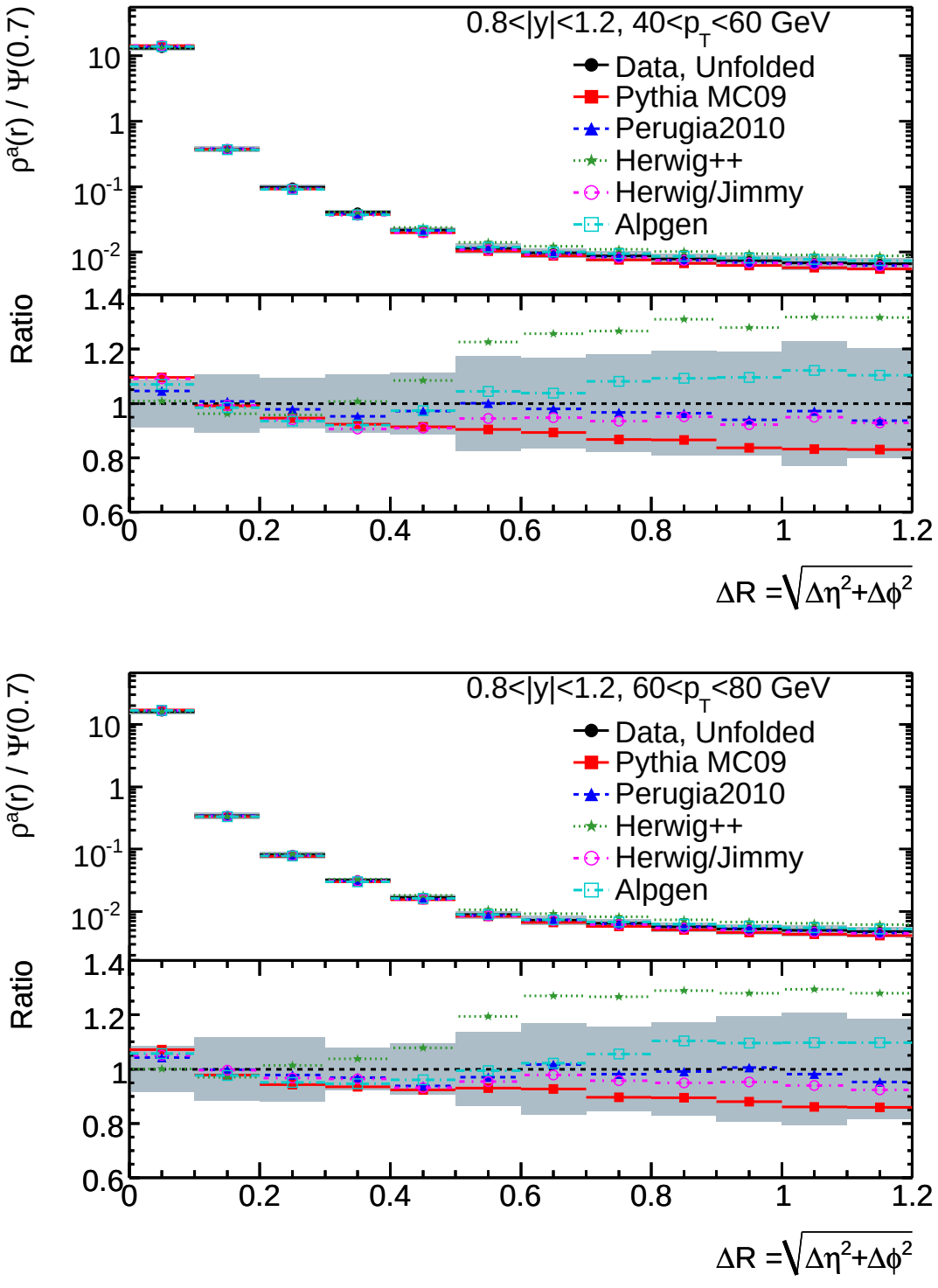


Figure A.26: Differential jet shapes after unfolding, for jets with $0.8 < |y| < 1.2$, in two p_T bins.

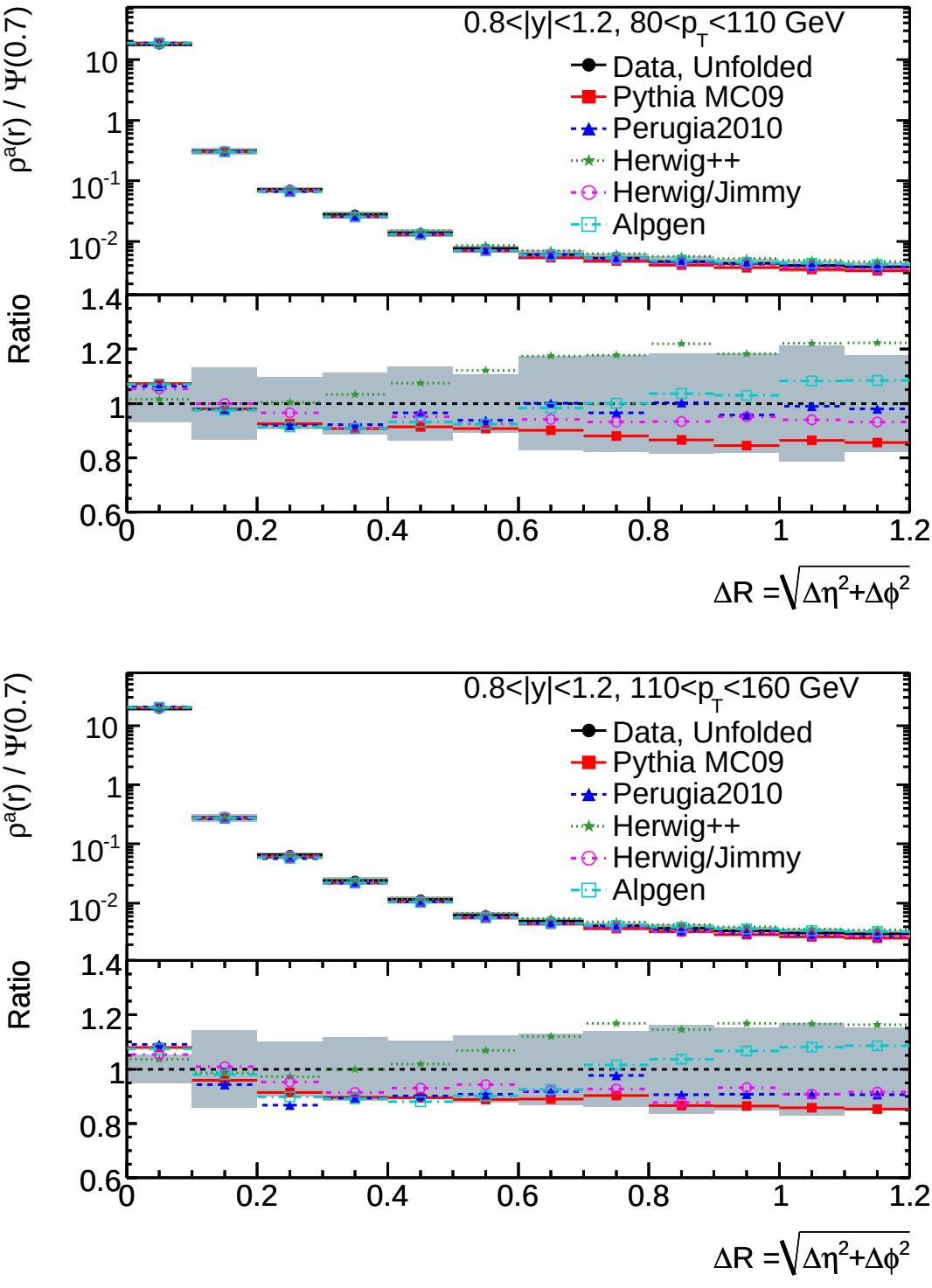


Figure A.27: Differential jet shapes after unfolding, for jets with $0.8 < |y| < 1.2$, in two p_T bins.

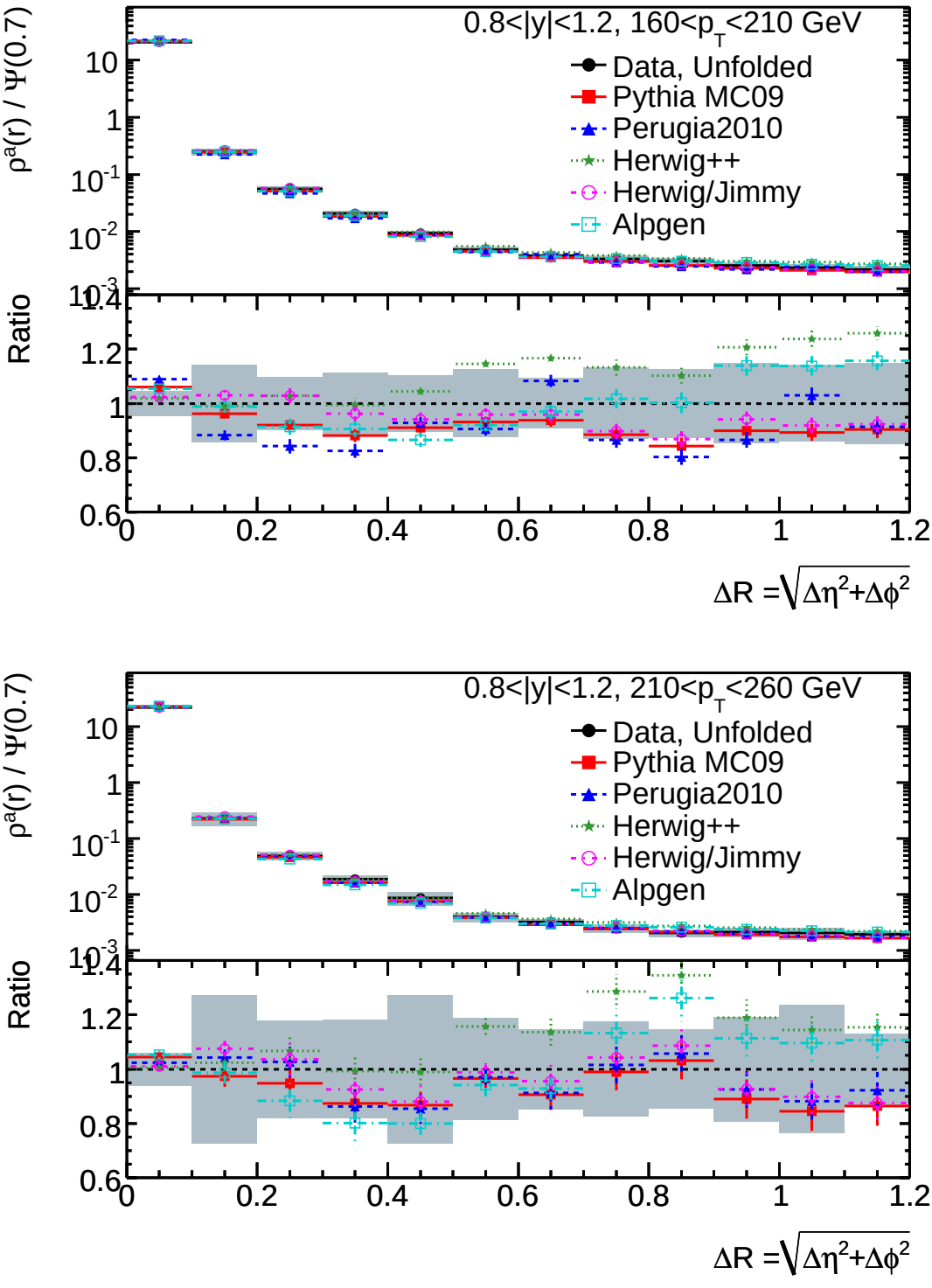


Figure A.28: Differential jet shapes after unfolding, for jets with $0.8 < |y| < 1.2$, in two p_T bins.

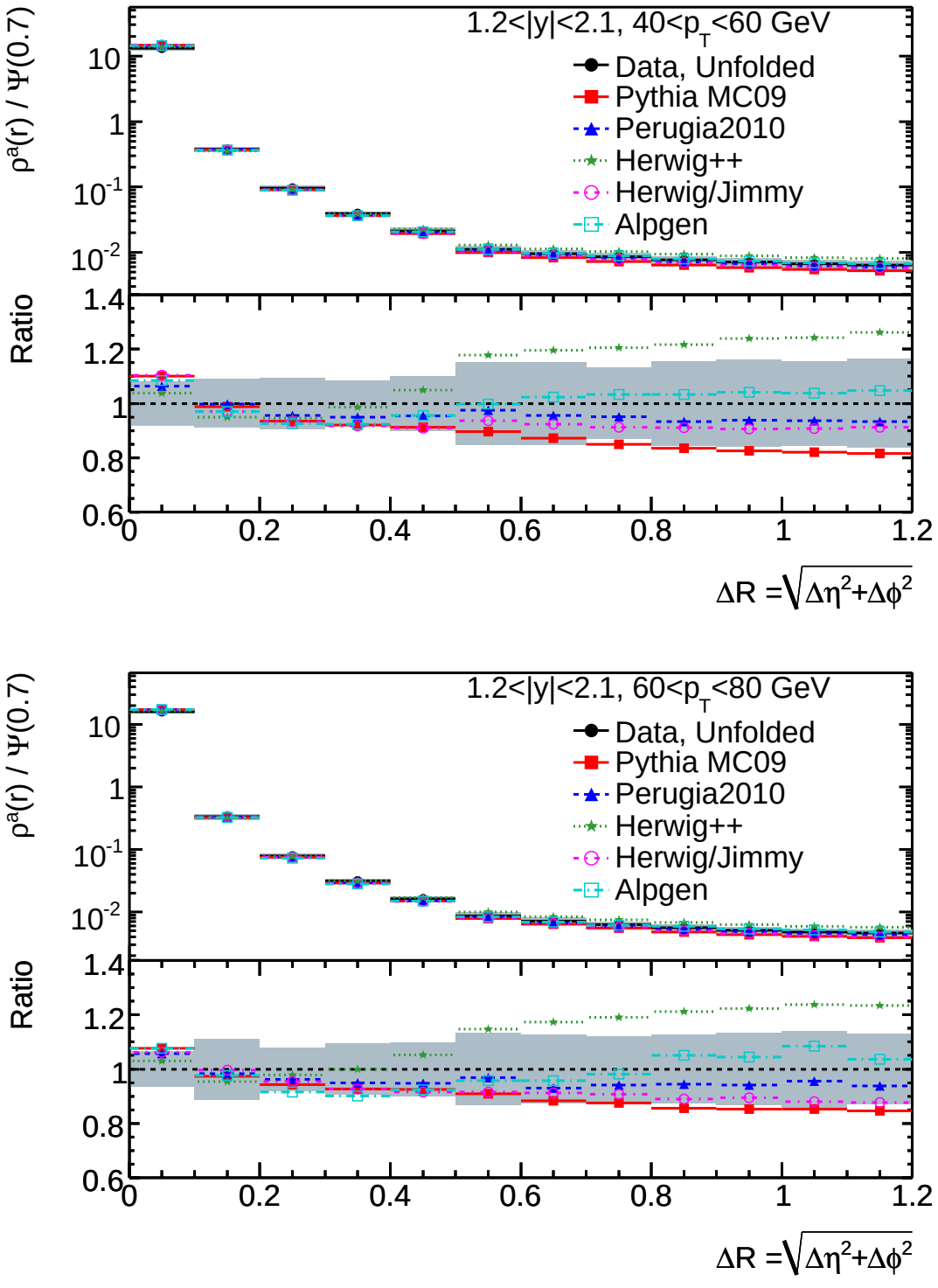


Figure A.29: Differential jet shapes after unfolding, for jets with $1.2 < |y| < 2.1$, in two p_T bins.

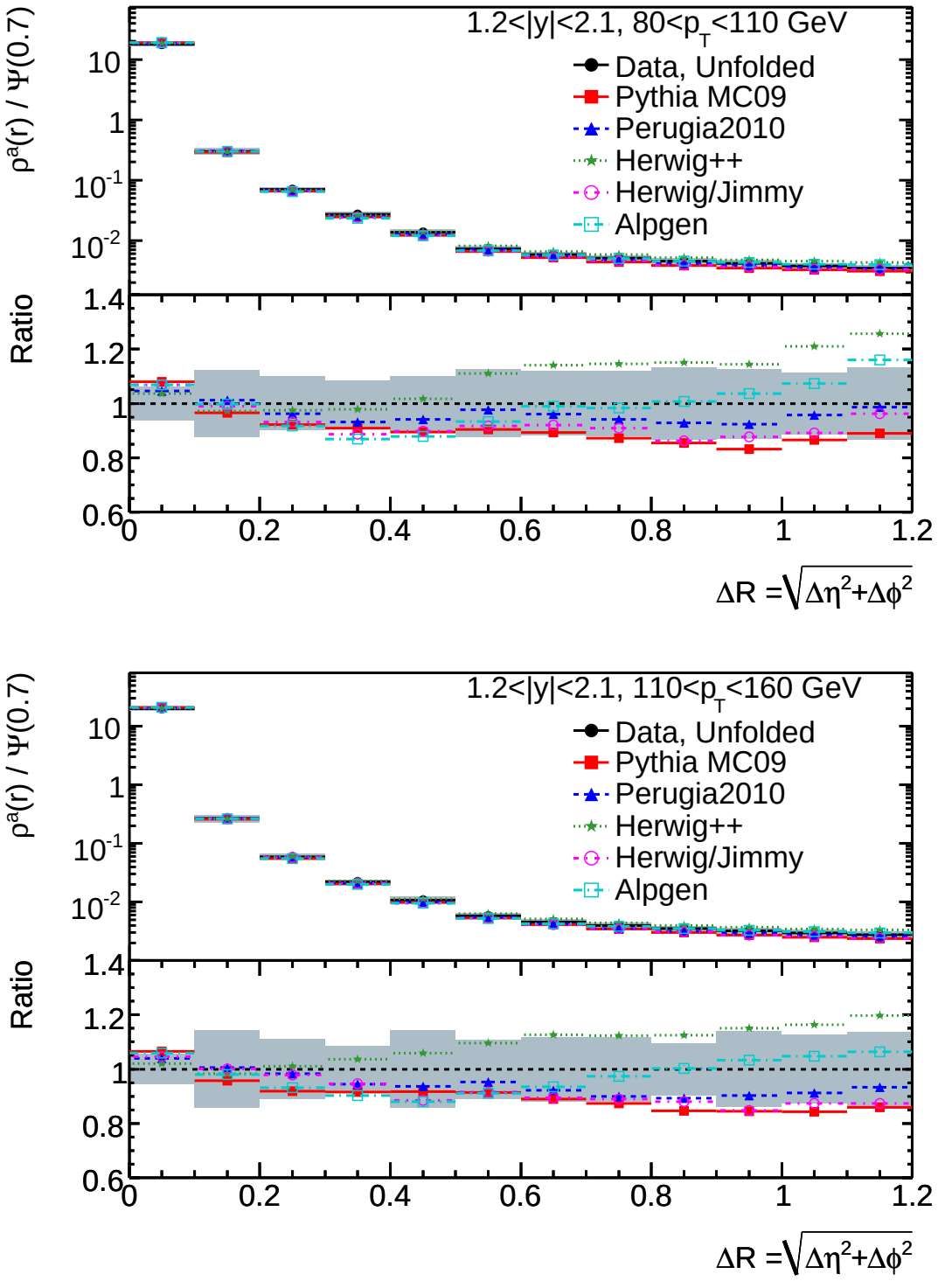


Figure A.30: Differential jet shapes after unfolding, for jets with $1.2 < |y| < 2.1$, in two p_T bins.

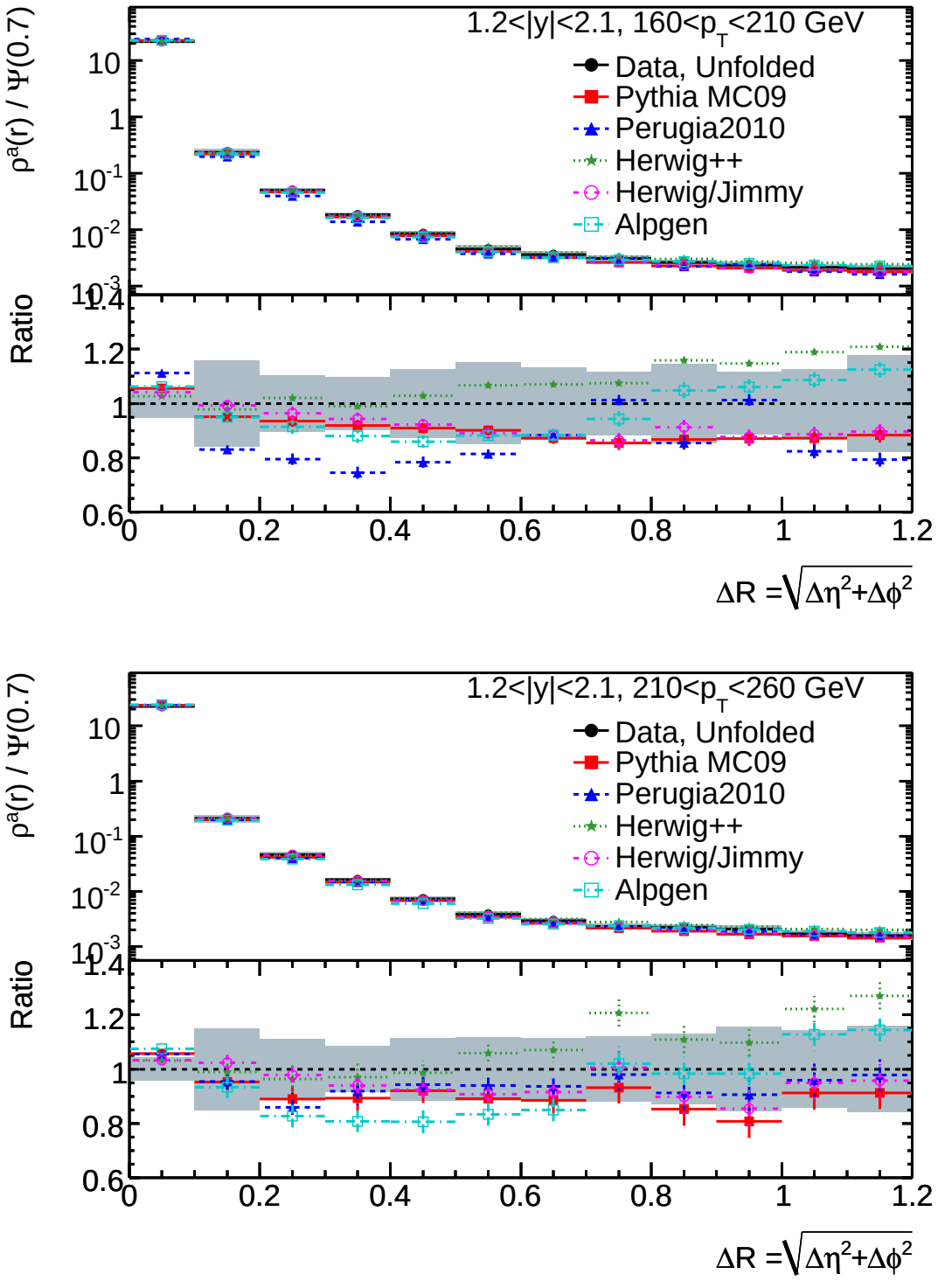


Figure A.31: Differential jet shapes after unfolding, for jets with $1.2 < |y| < 2.1$, in two p_T bins.

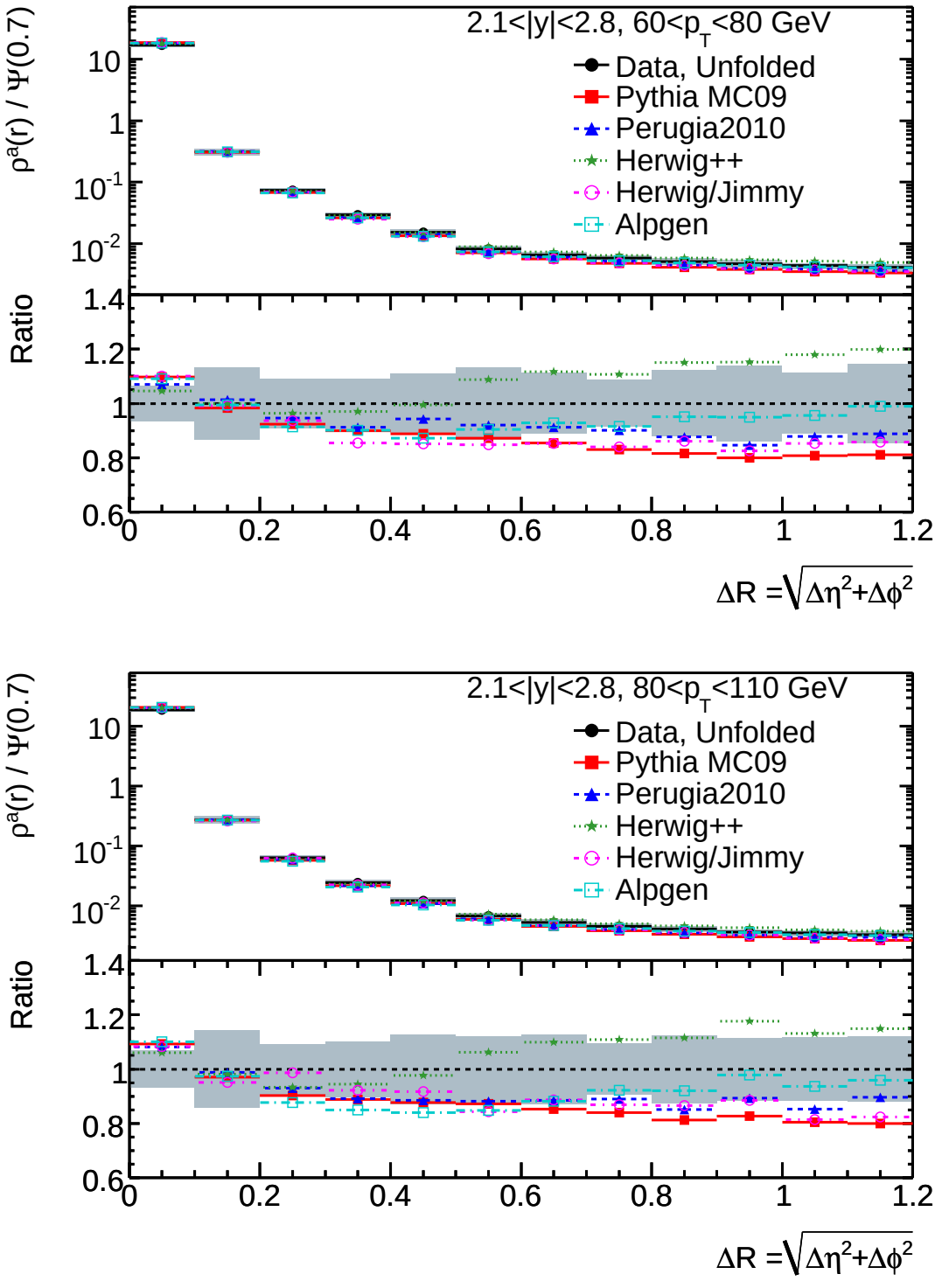


Figure A.32: Differential jet shapes after unfolding, for jets with $2.1 < |y| < 2.8$, in two p_T bins.

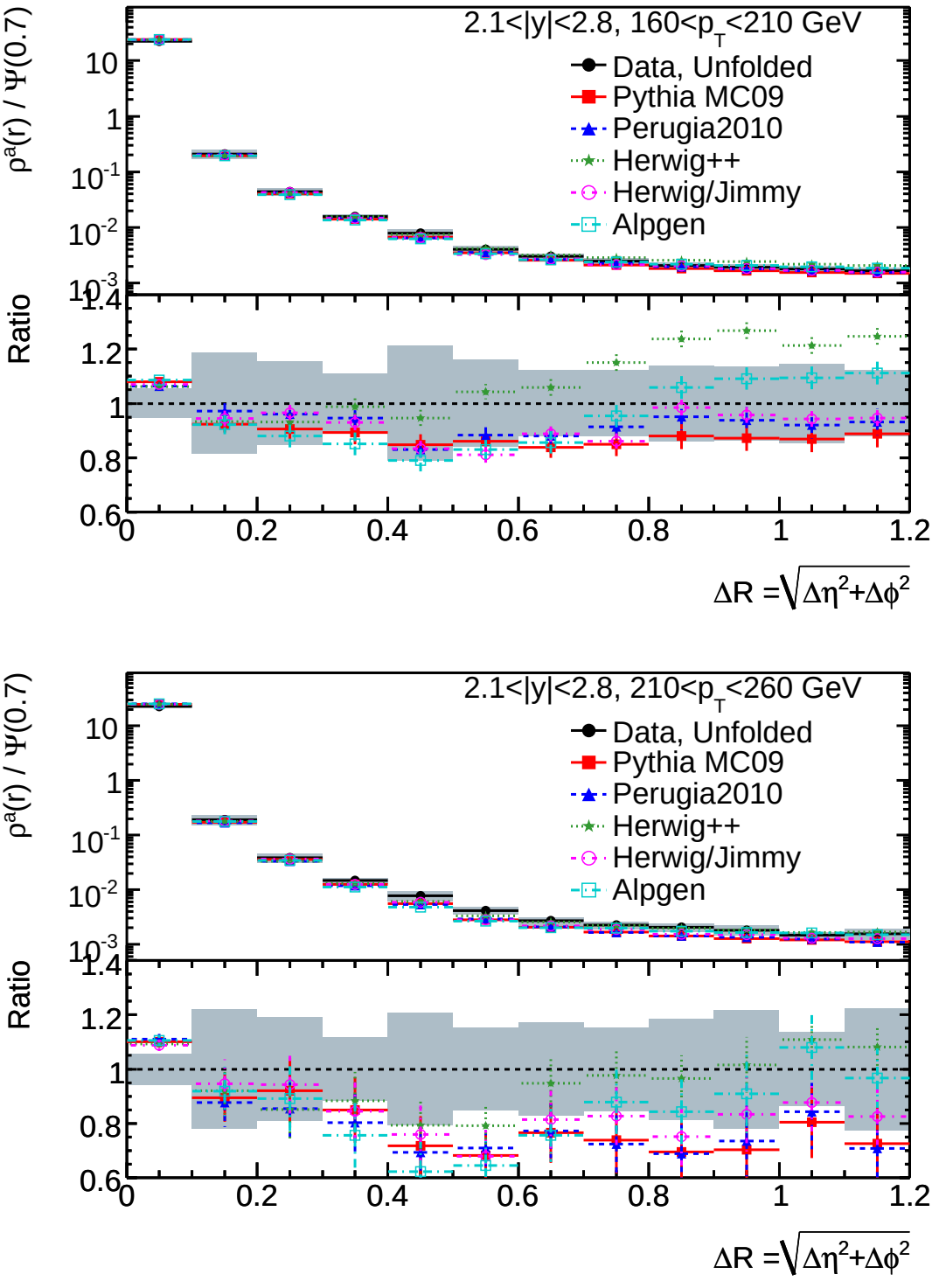


Figure A.33: Differential jet shapes after unfolding, for jets with $2.1 < |y| < 2.8$, in two p_T bins.

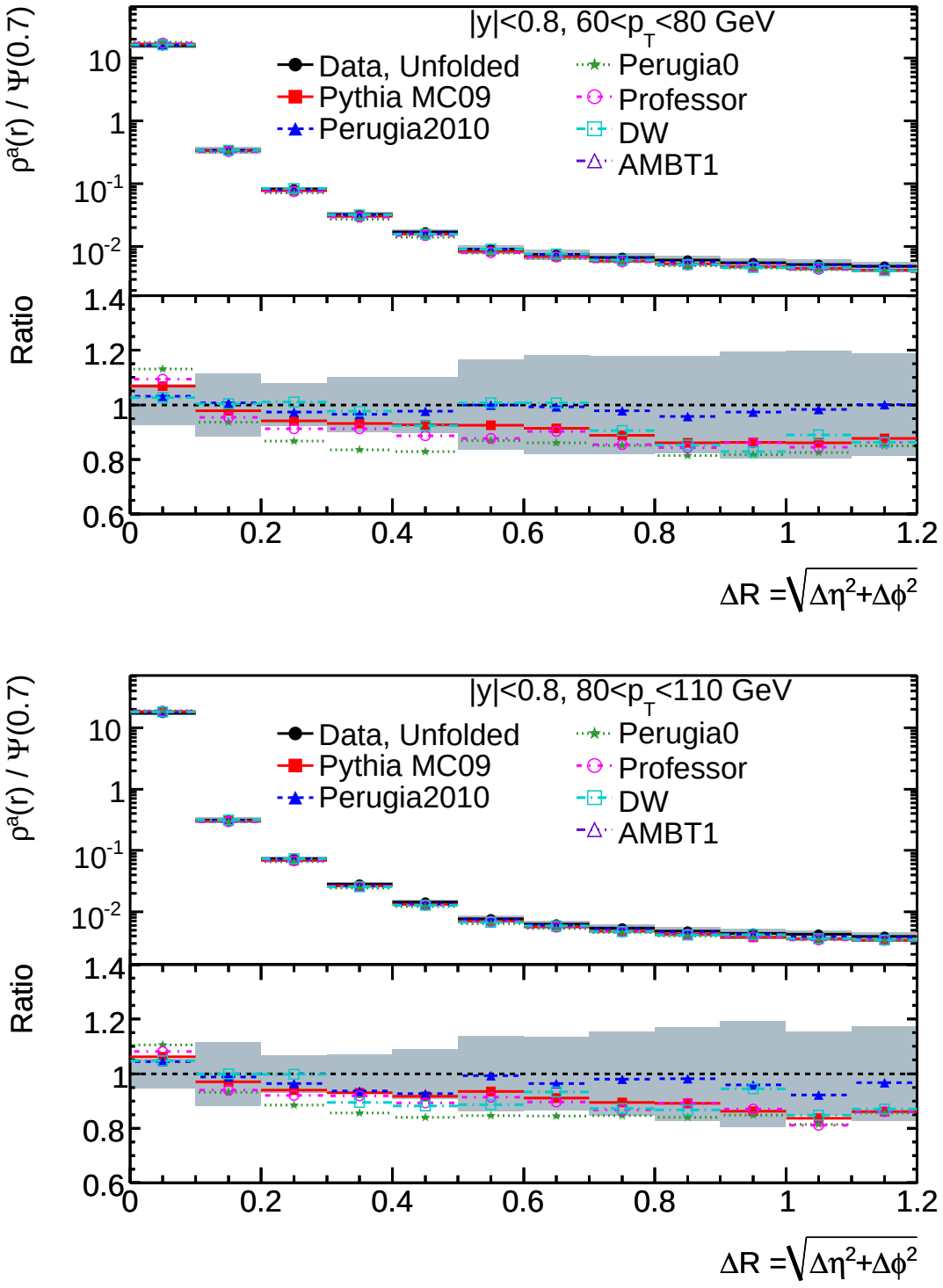


Figure A.34: Differential jet shapes after unfolding, for jets with $|y| < 0.8$, in two p_T bins.

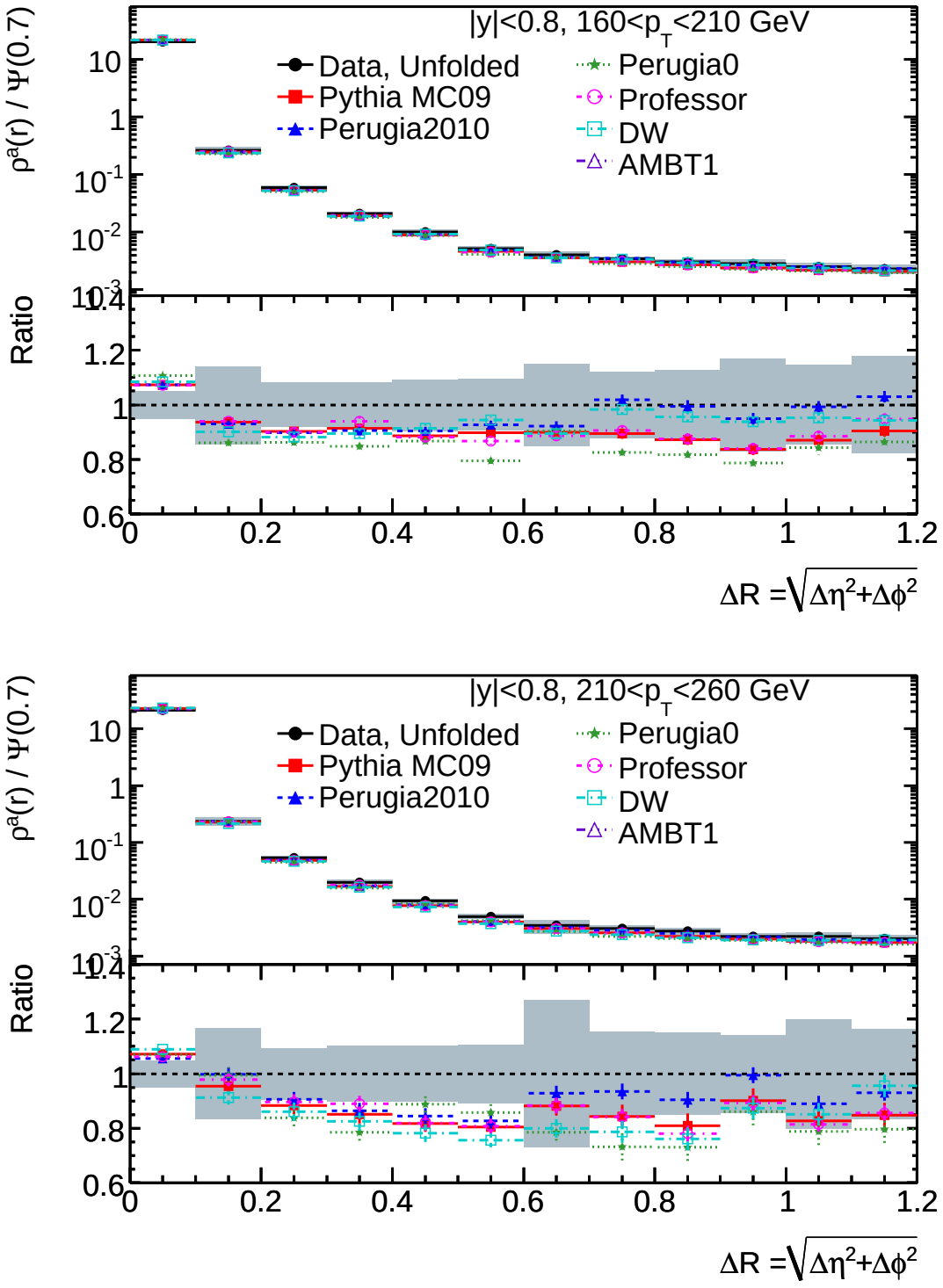


Figure A.35: Differential jet shapes after unfolding, for jets with $|y| < 0.8$, in two p_T bins.

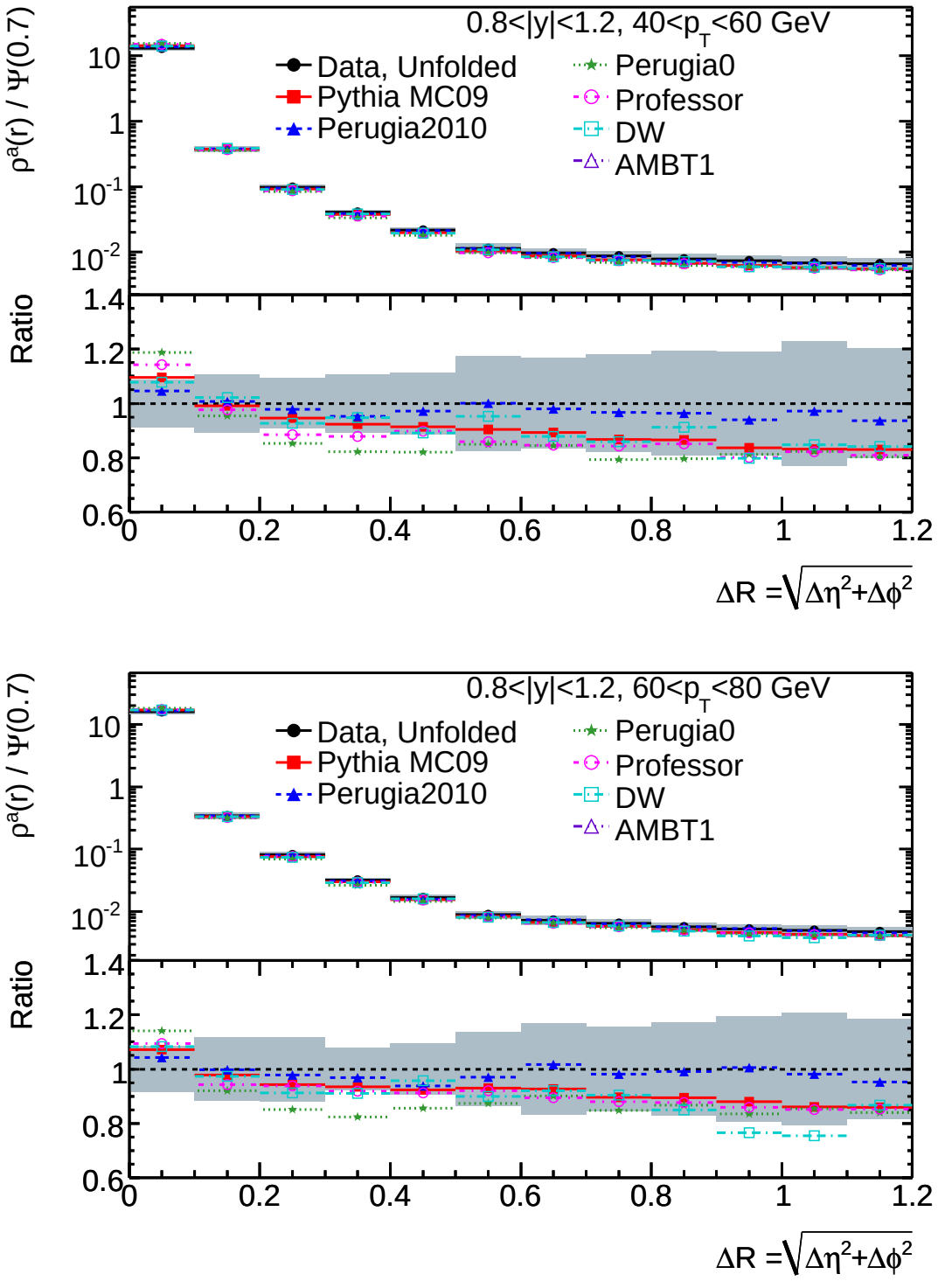


Figure A.36: Differential jet shapes after unfolding, for jets with $0.8 < |y| < 1.2$, in two p_T bins.

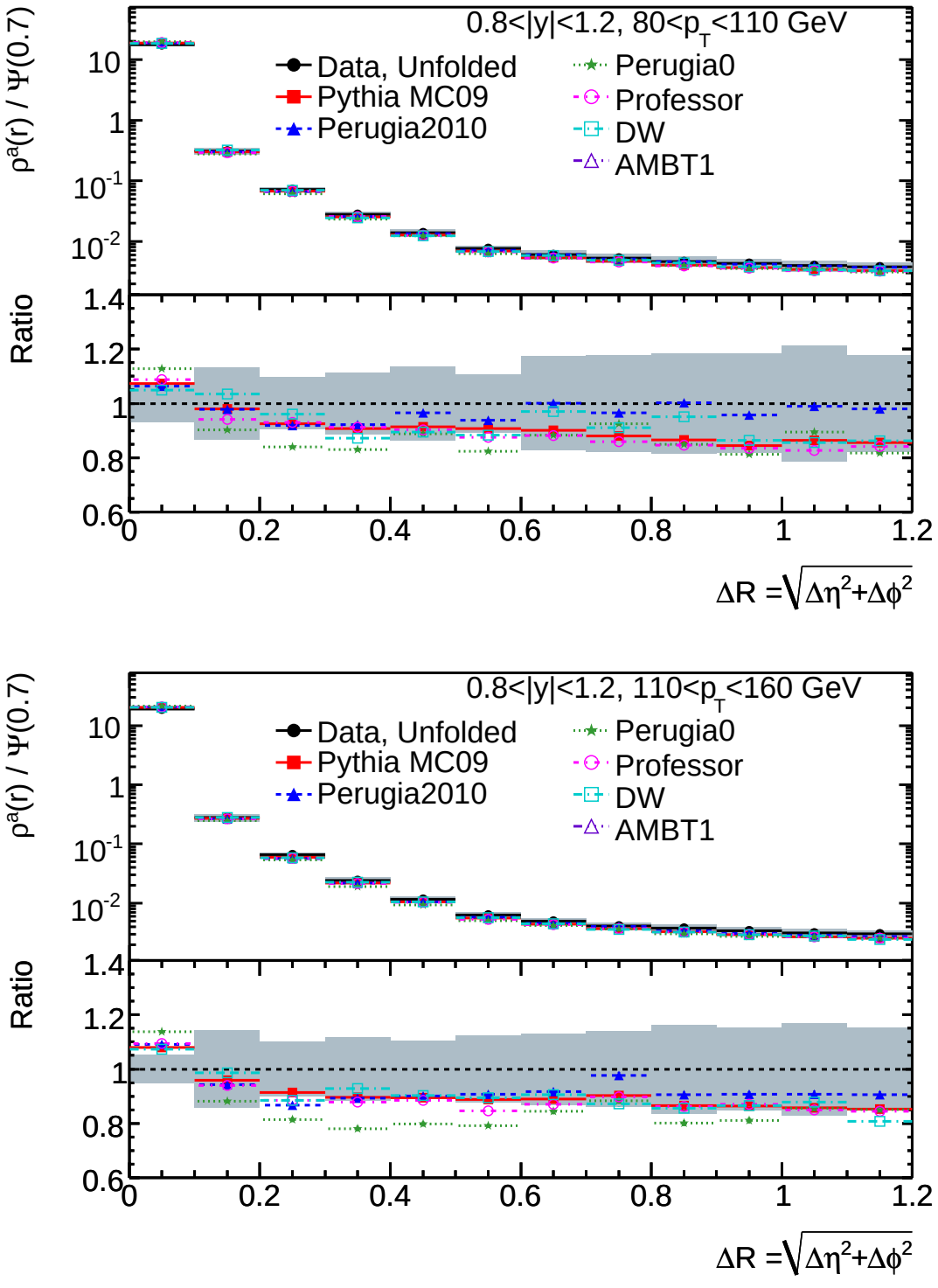


Figure A.37: Differential jet shapes after unfolding, for jets with $0.8 < |y| < 1.2$, in two p_T bins.

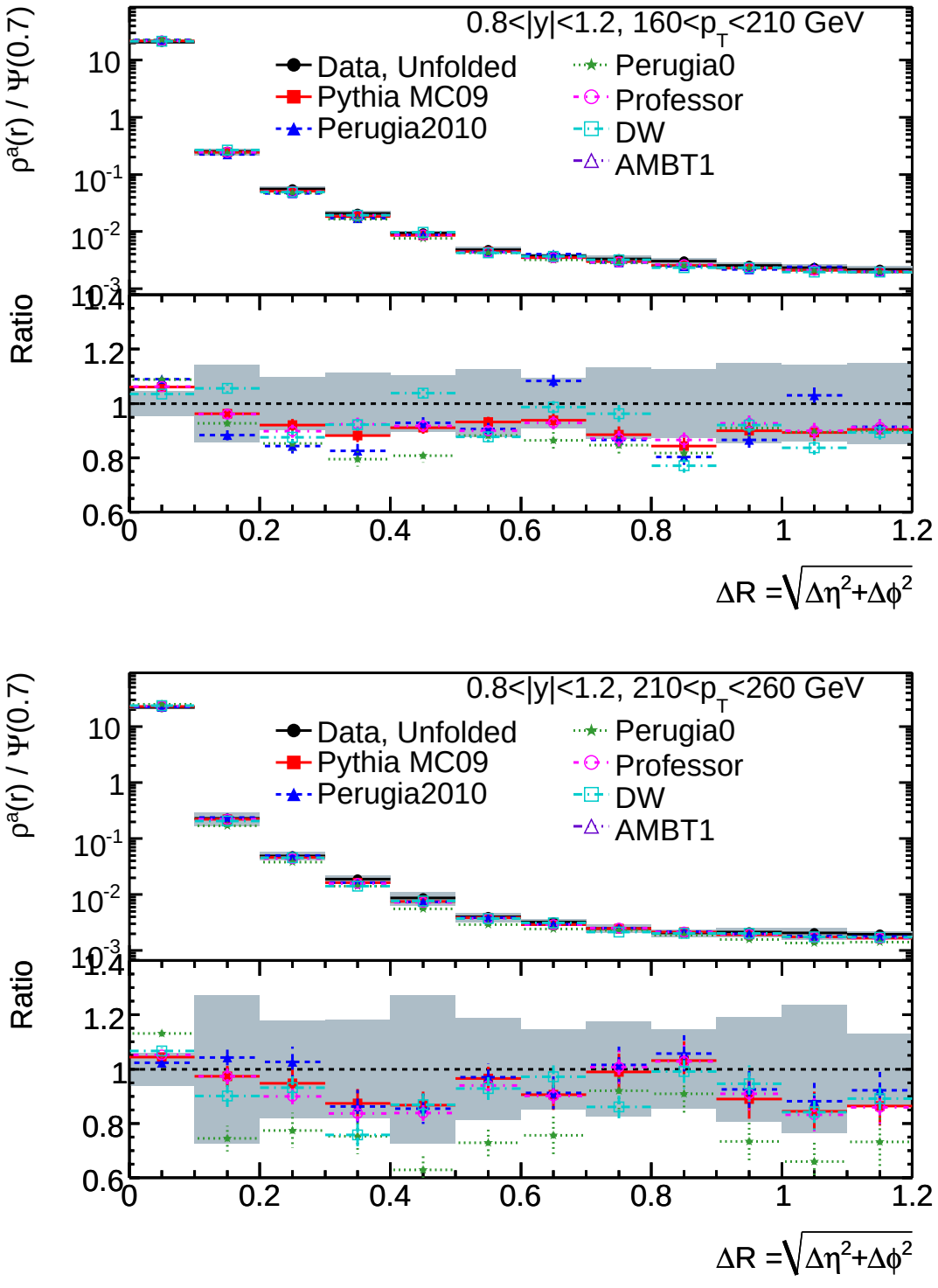


Figure A.38: Differential jet shapes after unfolding, for jets with $0.8 < |y| < 1.2$, in two p_T bins.

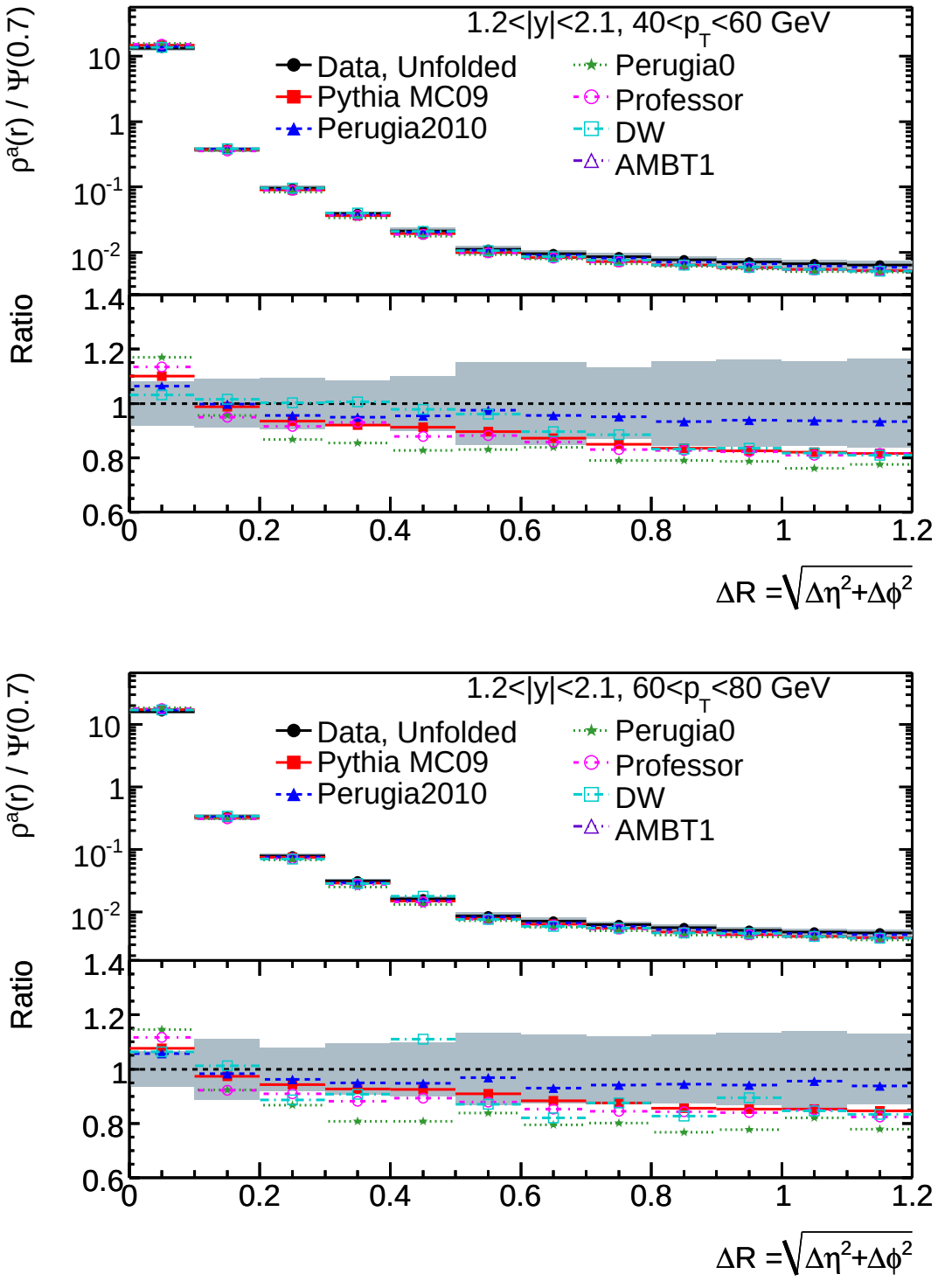


Figure A.39: Differential jet shapes after unfolding, for jets with $1.2 < |y| < 2.1$, in two p_T bins.

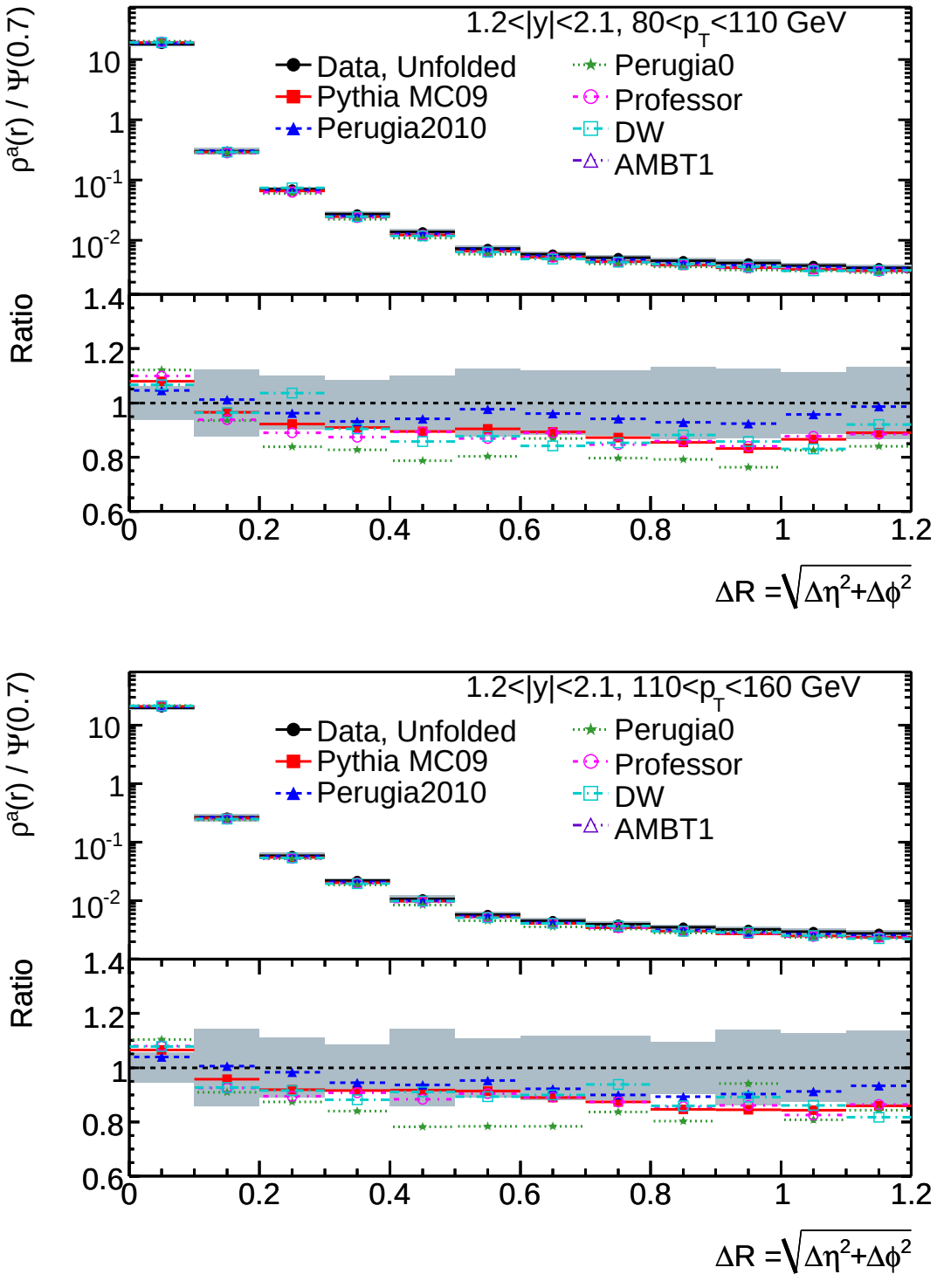


Figure A.40: Differential jet shapes after unfolding, for jets with $1.2 < |y| < 2.1$, in two p_T bins.

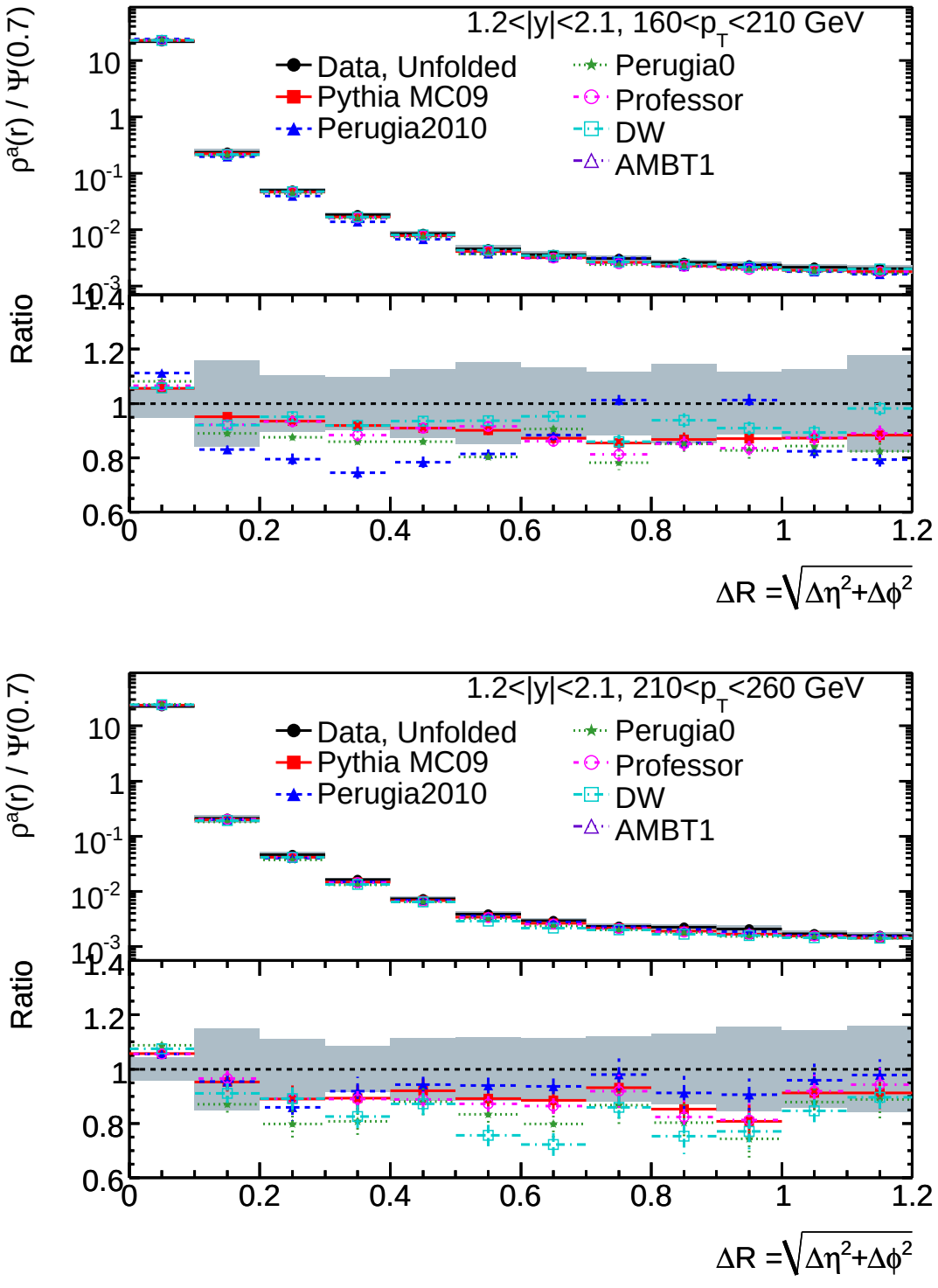


Figure A.41: Differential jet shapes after unfolding, for jets with $1.2 < |y| < 2.1$, in two p_T bins.

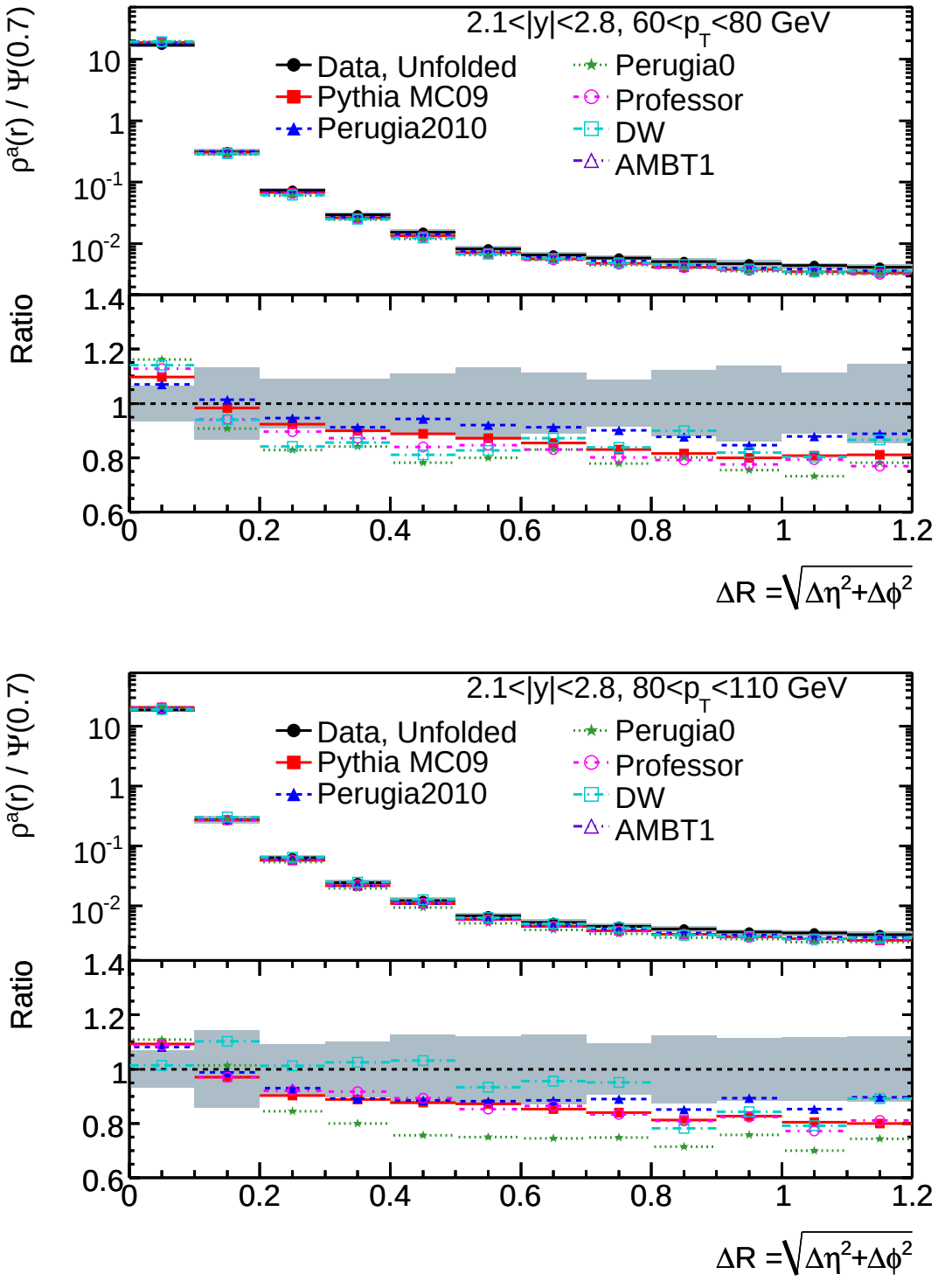


Figure A.42: Differential jet shapes after unfolding, for jets with $2.1 < |y| < 2.8$, in two p_T bins.

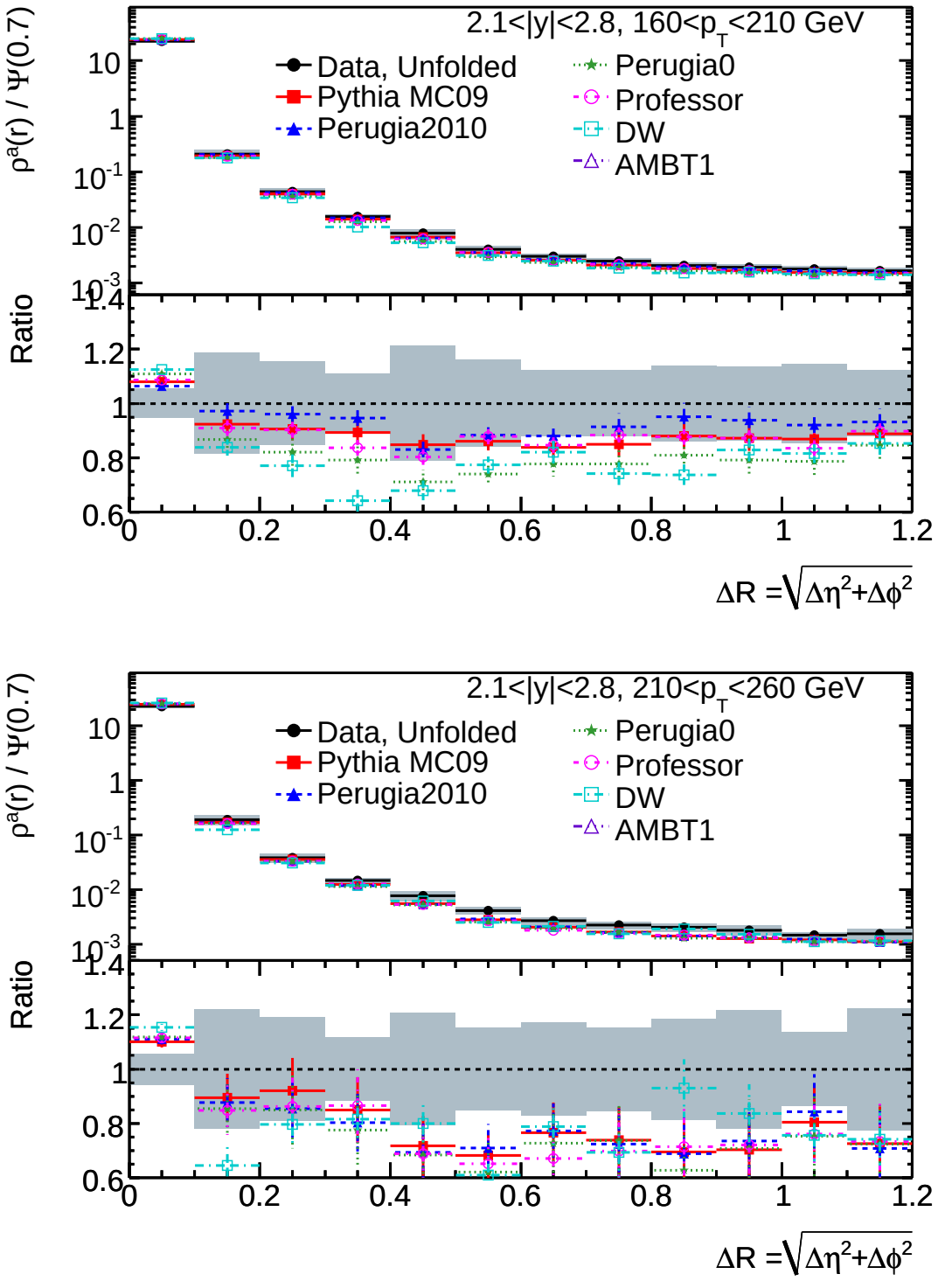


Figure A.43: Differential jet shapes after unfolding, for jets with $2.1 < |y| < 2.8$, in two p_T bins.

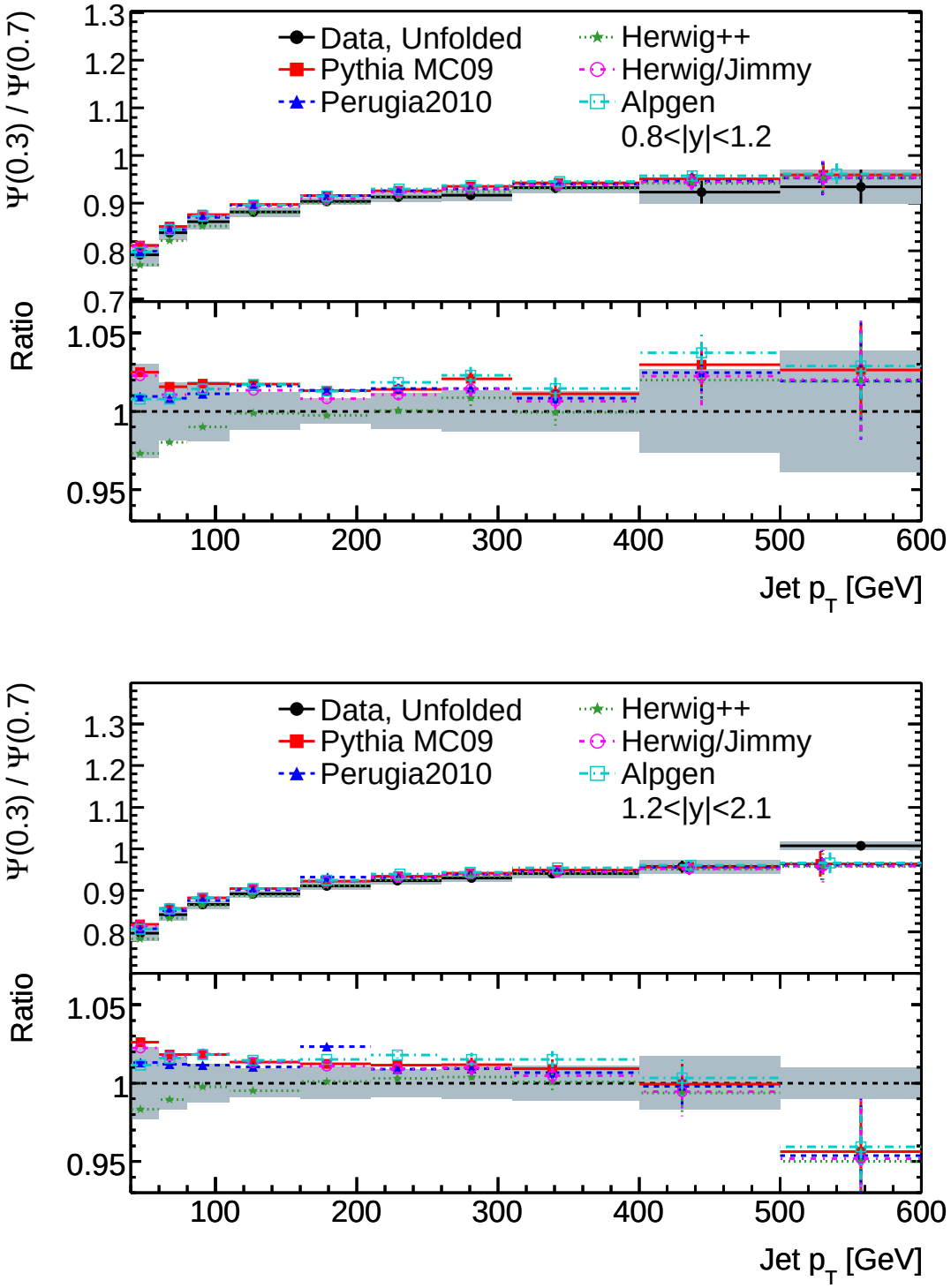


Figure A.44: Integrated jet shapes after unfolding as a function of jet p_T for jets with $0.8 < |y| < 1.1$ (top) and $1.2 < |y| < 2.1$ (bottom).

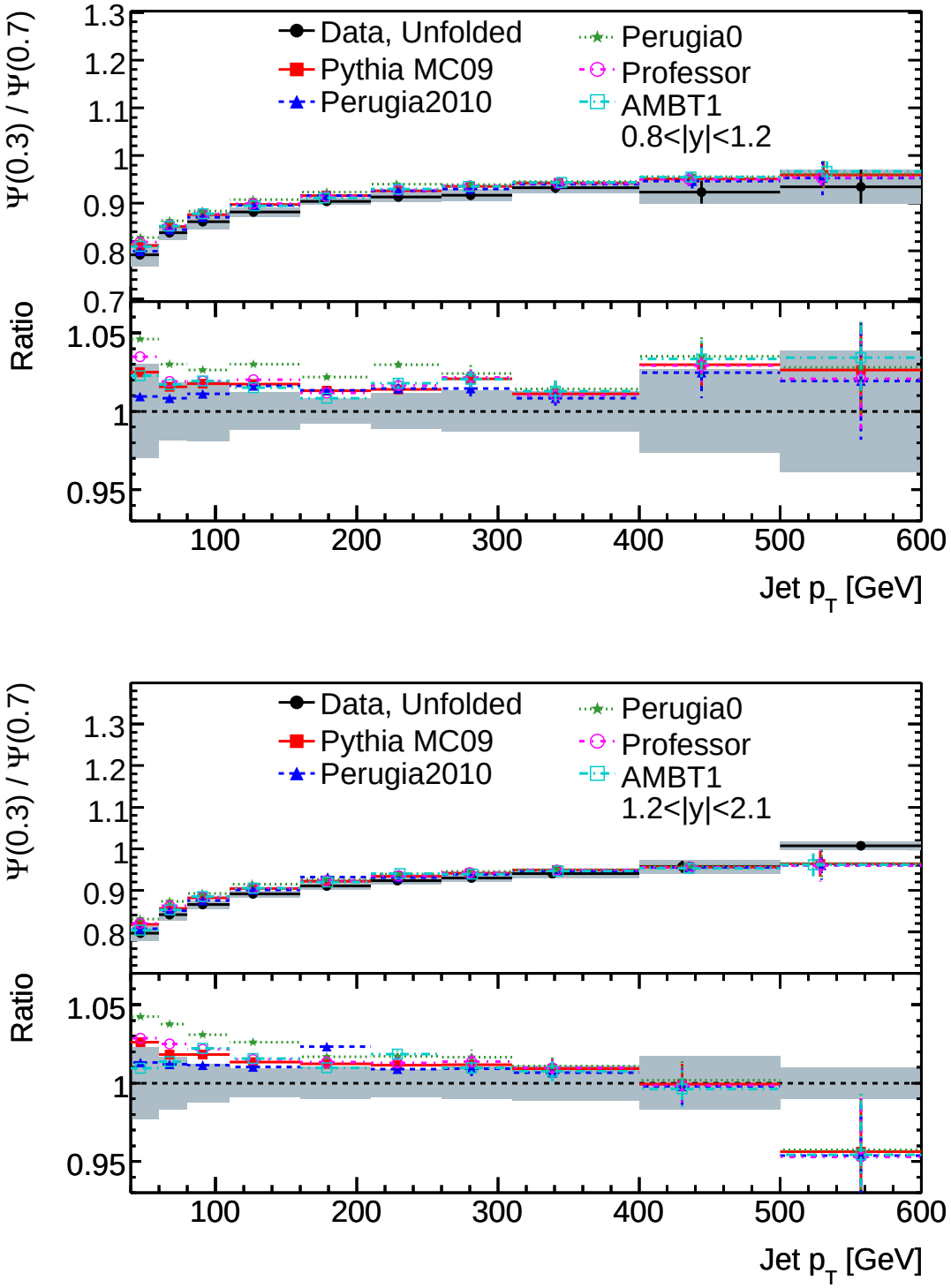


Figure A.45: Integrated jet shapes after unfolding as a function of jet p_T for jets with $0.8 < |y| < 1.1$ (top) and $1.2 < |y| < 2.1$ (bottom).

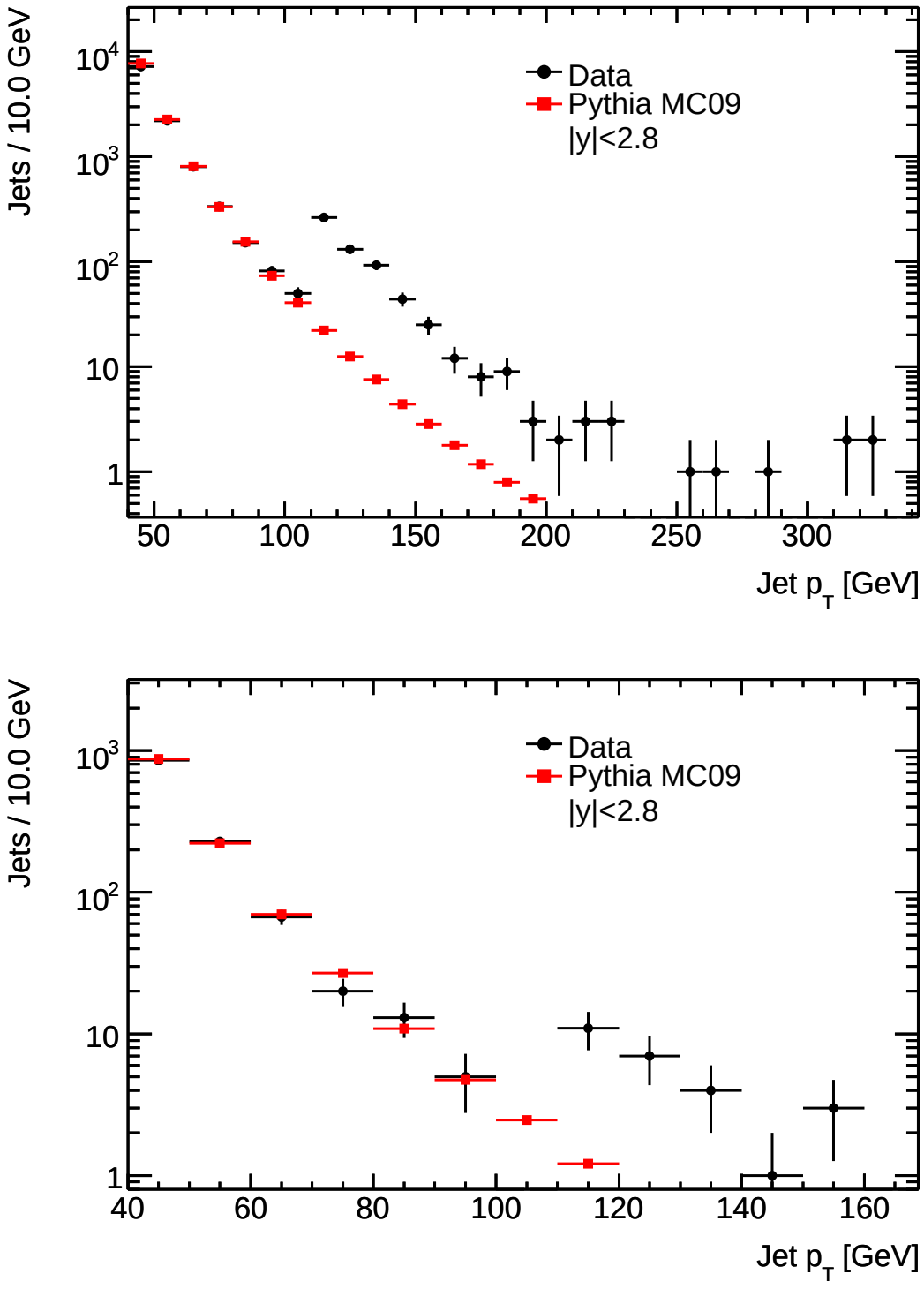


Figure A.46: Transverse momenta of the third and fourth leading jets in various generators at the hadron level. The discontinuity at 110 GeV is caused by the different triggers being used above and below that p_T.

Bibliography

[1] M. Gell-Mann, M.L. Goldberger, N.M. Kroll, and F.E. Low, *Amelioration of Divergence Difficulties in the Theory of Weak Interactions*. Phys. Rev. **179**, (1969) 1518–1527, [doi: 10.1103/PhysRev.179.1518].

[2] P.A.M. Dirac, *The Quantum Theory of the Electron*. Proc. R. Soc. Lond. **A 117**, (1928) 610–624, [doi: 10.1098/rspa.1928.0023].

[3] H. Yukawa, *On the Interaction of Elementary Particles*. Proc. Phys. Math. Soc. Jap. **17**, (1935) 48–57.

[4] C.M.G. Lattes, H. Muirhead, G.P.S. Occhialini, and C.F. Powell, *Processes Involving Charged Mesons*. Nature **159**, (1947) 694–697, [doi: 10.1038/159694a0].

[5] C.M.G. Lattes, G.P.S. Occhialini, and C.F. Powell, *Observation on the Tracks of Slow Mesons in Photographic Emulsions*. Nature **160**, (1947) 453–456, [doi: 10.1038/160453a0].

[6] C.M.G. Lattes, G.P.S. Occhialini, and C.F. Powell, *Observation on the Tracks of Slow Mesons in Photographic Emulsions*. Nature **160**, (1947) 486–492, [doi: 10.1038/160486a0].

[7] C.D. Anderson, *The apparent existence of easily deflectable positives*. Science **76**, (1932) 238–239, [doi: 10.1126/science.76.1967.238].

[8] C.D. Anderson, *The positive electron*. Phys. Rev. **43**, (1933) 491–494, [doi: 10.1103/PhysRev.43.491].

[9] G.D. Rochester and C.C. Butler, *Evidence for the Existence of New Unstable Elementary Particles*. Nature **160**, (1947) 855, [doi: 10.1038/160855a0].

[10] V.V. Ezhela et al., *Particle Physics: One Hundred Years of Discoveries: An Annotated Chronological Bibliography* (American Institute of Physics, Woodbury, NY) (1996), [ISBN: 9781563966422].

[11] M. Gell-Mann and Y. Ne'eman, *The Eightfold Way* (W.A. Benjamin, New York) (1964), [ISBN: 9780738202990].

[12] M. Gell-Mann, *A schematic model of baryons and mesons*. Phys. Lett. **8**, (1964) 214–215, [doi: 10.1016/S0031-9163(64)92001-3].

[13] G. Zweig, *An SU(3) model for strong interaction symmetry and its breaking*. CERN-TH 401 (1964).

[14] G. Zweig, *An SU(3) model for strong interaction symmetry and its breaking 2*. CERN-TH 412 (1964).

[15] E.D. Bloom et al., *High-Energy Inelastic e-p Scattering at 6° and 10°*. Phys. Rev. Lett. **23**, (1969) 930–934, [doi: 10.1103/PhysRevLett.23.930].

[16] M. Breidenbach et al., *Observed Behavior of Highly Inelastic Electron-Proton Scattering*. Phys. Rev. Lett. **23**, (1969) 935–939, [doi: 10.1103/PhysRevLett.23.935].

[17] F. Abe et al., the **CDF** Collaboration, *Observation of Top Quark production in $p\bar{p}$ Collisions with the Collider Detector at Fermilab*. Phys. Rev. Lett. **74**, (1995) 2626–2631, [doi: 10.1103/PhysRevLett.74.2626], [hep-ex/9503002].

[18] S. Abachi et al., the **DZero** Collaboration, *Search for High Mass Top Quark Production in $p\bar{p}$ Collisions at $\sqrt{s}$=1.8 TeV*. Phys. Rev. Lett. **74**, (1995) 2422–2426, [doi: 10.1103/PhysRevLett.74.2422], [hep-ex/9411001].

[19] P.F. Smith, *Searches for fractional electric charge in terrestrial materials*. Ann. Rev. Nucl. Part. Sci. **39**, (1989) 73–111, [doi: 10.1146/annurev.ns.39.120189.000445].

[20] M.L. Perl, E.R. Lee, and D. Loomba, *Searches for fractionally charged particles*. Ann. Rev. Nucl. Part. Sci. **59**, (2009) 47–65, [doi: 10.1146/annurev-nucl-121908-122035].

[21] O.W. Greenberg, *Spin and Unitarity-Spin Independence in a Paraquark Model of Baryons and Mesons*. Phys. Rev. Lett. **13**, (1964) 598–602, [doi: 10.1103/PhysRevLett.13.598].

[22] M.Y. Han and Y. Nambu, *Three-Triplet Model with Double SU(3) Symmetry*. Phys. Rev. **139**, (1965) B1006–B1010, [doi: 10.1103/PhysRev.139.B1006].

[23] P.G.O. Freund and Y. Nambu, *Dynamics of the Zweig-Iizuka Rule and a New Vector Meson below 2 GeV/c^2*. Phys. Rev. Lett. **34**, (1975) 1645–1649, [doi: 10.1103/PhysRevLett.34.1645].

[24] S. Meshkov, *Glueballs*. In *Proceedings of the 7th International Conference on Experimental Meson Spectroscopy, Upton, NY, 14-16 April 1983*, edited by S.J. Lindenbaum (American Institute of Physics, New York) (1983), [ISBN: 9780883183120], AIP Conference Proceedings, 113.

[25] W. Bietenholz et al., *Exploring the Nucleon Structure from First Principles of QCD*. In *Proceedings of the XXXIII Symposium on Nuclear Physics, Cocoyoc (Morelos, Mexico), January 5-8, 2010*, edited by Libertad Barrón-Palos, Roelof Bijker, Ruben Fossion, and David Lizcano, 012011 (IOP Journal of Physics: Conference Series) (2010), [doi: 10.1088/1742-6596/239/1/012011], [arXiv:1004.2100].

[26] J.D. Bjorken, *Asymptotic Sum Rules at Infinite Momentum*. Phys. Rev. **179**, (1969) 1547–1553, [doi: 10.1103/PhysRev.179.1547].

[27] R.P. Feynman, *The Behavior of Hadron Collisions at Extreme Energies*. In *Proceedings of the Third International Conference on High Energy Collision, Stony Brook, NY, 5-6 September 1969*, edited by C.N. Yang et al., 237–249 (Gordon and Breach, New York) (1969).

[28] H.D. Politzer, *Reliable perturbative results for strong interactions*. Phys. Rev. Lett. **30**, (1973) 1346–1349, [doi: 10.1103/PhysRevLett.30.1346].

[29] D.J. Gross and F. Wilczek, *Ultraviolet behavior of non-Abelian gauge theories*. Phys. Rev. Lett. **30**, (1973) 1343–1346, [doi: 10.1103/PhysRevLett.30.1343].

[30] S.L. Glashow, *Partial-symmetries of weak interactions*. Nucl. Phys. **22**, (1961) 579–588, [doi: 10.1016/0029-5582(61)90469-2].

[31] S. Weinberg, *A Model of Leptons*. Phys. Rev. Lett. **19**, (1967) 1264–1266, [doi: 10.1103/PhysRevLett.19.1264].

[32] G. van't Hooft, *Renormalizable Lagrangians for massive Yang-Mills fields*. Nucl. Phs. **B35**, (1971) 167–188, [doi: 10.1016/0550-3213(71)90139-8].

[33] A. Salam, *Weak and electromagnetic interactions*. In *Proceedings of the 8th Nobel Symposium - Elementary Particle Theory: relativistic groups and analyticity, Aspenäsgøarden, Sweden, 19-25 May 1968*, edited by N. Svartholm, 367–372 (Almquist and Wiksells, Stockholm) (1969).

[34] A. Salam and J.C. Ward, *Electromagnetic and weak interactions*. Phys. Lett. **13**, (1964) 168–171, [doi: 10.1016/0031-9163(64)90711-5].

[35] S. Weinberg, *Non-Abelian gauge theories of strong interactions*. Phys. Rev. Lett. **31**, (1973) 494–497, [doi: 10.1103/PhysRevLett.31.494].

[36] A.D. Martin, W.J. Stirling, R.S. Thorne, and G. Watt, *Parton distributions for the LHC*. Eur. Phys. J. **C63**, (2009) 189–285, [doi: 10.1140/epjc/s10052-009-1072-5], [arXiv:0901.0002], figures from the MSTW Website.

[37] A. Sherstnev and R.S. Thorne, *Parton Distributions for LO Generators*. Eur. Phys. J. **C55**, (2008) 553–575, [doi: 10.1140/epjc/s10052-008-0610-x], [arXiv:0711.2473].

[38] W.K. Tung et al., *Global QCD Analysis and Collider Phenomenology–CTEQ*. In *Proceedings of the 15th International Workshop on Deep-Inelastic Scattering and Related Subjects, Munich, April 2007*, edited by G. Grindhammer and K. Sachs (Atlantis Press, Amsterdam) (2007), [ISBN: 9783935702232], [doi: 10.3360/dis.2007.28], [arXiv:0707.0275].

[39] D. Bourilkov, R.C. Group, and M.R. Whalley, *LHAPDF: PDF use from the Tevatron to the LHC*. In *Proceedings of the TeV4LHC Workshop - 4th meeting, Batavia, Illinois, 20-22 October 2005* (2006), [hep-ph/0605240].

[40] R.K. Ellis, W.J. Stirling, and B.R. Webber, *QCD and Collider Physics* (Cambridge University Press, Cabridge) (1996), [ISBN: 9780521545891], [doi: 10.2277/0521545897].

[41] C. Buttar et al., *The QCD, EW, and Higgs Working Group: Summary Report*. In *Proceedings of the Fourth Workshop on Physics at TeV Colliders, Les Houches, France, 2-20 May 2005*, edited by G. Bélanger, F. Boudjema, J.Ph. Guillet, and E. Pilon (2005), [hep-ph/0604120v1].

[42] M.L. Mangano et al., *ALPGEN, a generator for hard multiparton processes in hadronic collisions*. JHEP **07**, (2003) 001, [doi: 10.1088/1126-6708/2003/07/001], [hep-ph/0206293].

[43] M.L. Mangano and S.J. Parke, *Multi-parton amplitudes in gauge theories*. Physics Reports **200**, (1991) 301–367, [doi: 10.1016/0370-1573(91)90091-Y].

[44] T. Sjöstrand, *Monte Carlo Tools*. In *Proceedings of the 65th Scottish Universities Summer School in Physics: LHC Physics, St. Andrews, 16-29 August 2009*, edited by T. Binoth, C. Buttar, and P.J. Clark (Taylor and Francis) (2011), [ISBN: 978-1-4398377-0-2], to be published, [arXiv:0911.5286].

[45] T.T. Chou and C.N. Yang, *Model of Elastic High-Energy Scattering*. Phys. Rev. **170**, (1968) 1591–1596, [doi: 10.1103/PhysRev.170.1591].

[46] D.E. Acosta et al., the **CDF** Collaboration, *The underlying event in hard interactions at the Tevatron $\bar{p}p$ collider*. Phys. Rev. **D70**, (2004) 072002, [doi: 10.1103/PhysRevD.70.072002], [hep-ex/0404004].

[47] A. Moraes, *Modeling the underlying event: generating predictions for the LHC*. In *Proceedings of the First International Workshop on Multiple Partonic Interactions at the LHC (MPI@LHC08), Perugia, Italy, October 27-31, 2008*, edited by P. Bartalini and L. Fanò, 124–130 (DESY-PROC-2009-06, DESY, Germany) (2010), [arXiv:1003.4220].

[48] S. Catani, F. Krauss, R. Kuhn, and B.R. Webber, *QCD Matrix Elements + Parton Showers*. JHEP **0111**, (2001) 063, [doi: 10.1088/1126-6708/2001/11/063], [hep-ph/0109231].

[49] L. Lönnblad, *Correcting the Colour-Dipole Cascade Model with Fixed Order Matrix Elements*. JHEP **0205**, (2002) 046, [doi: 10.1088/1126-6708/2002/05/046].

[50] M.L. Mangano, M. Moretti, F. Piccini, and M. Treccani, *Matching matrix elements and shower evolution for top-pair production in hadronic collisions*. JHEP **0701**, (2007) 013, [doi: 10.1088/1126-6708/2007/01/013], [hep-ph/0611129].

[51] V.N. Gribov and L.N. Lipatov, *Deep inelastic e p scattering in perturbation theory*. Sov. J. Nucl. Phys. **15**, (1972) 438–450.

[52] G. Altarelli and G. Parisi, *Asymptotic freedom in parton language*. Nucl. Phys. **B126**, (1977) 298, [doi: 10.1016/0550-3213(77)90384-4].

[53] Yu.L. Dokshitzer, *Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics*. Sov. Phys. JETP **46**, (1977) 641–653.

[54] T. Sjöstrand, S. Mrenna, and P. Skands, *PYTHIA 6.4 physics and manual*. JHEP **05**, (2006) 026, [doi: 10.1088/1126-6708/2006/05/026], [hep-ph/0603175].

[55] T. Gleisberg et al., *SHERPA 1.α, a proof-of-concept version*. JHEP **02**, (2004) 056, [doi: 10.1088/1126-6708/2004/02/056], [hep-ph/0311263].

[56] G. Marchesini et al., *HERWIG: A Monte Carlo event generator for simulating hadron emission reactions with interfering gluons. Version 5.1 - April 1991*. Comput. Phys. Commun. **67**, (1992) 465, [doi: 10.1016/0010-4655(92)90055-4].

[57] G. Corcella et al., *HERWIG 6: An event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)*. JHEP **01**, (2001) 010, [doi: 10.1088/1126-6708/2001/01/010], [hep-ph/0011363].

[58] M. Bahr et al., *Herwig++ Physics and Manual*. Eur. Phys. J. **C58**, (2008) 639–707, [doi: 10.1140/epjc/s10052-008-0798-9], [arXiv:0803.0883].

[59] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand, *Parton fragmentation and string dynamics*. Phys. Rep. **97**, (1983) 31, [doi: 10.1016/0370-1573(83)90080-7].

[60] B.R. Webber, *A QCD Model for Jet Fragmentation Including Soft Gluon Interference*. Nucl. Phys. **B238**, (1984) 492–528, [doi: 10.1016/0550-3213(84)90333-X].

[61] C.M. Buttar et al., *The Underlying Event*. In *HERA and the LHC: A workshop on the implications of HERA for LHC physics*, edited by A. de Roeck and H. Jung (DESY-PROC-2005-001, DESY, Germany) (2005), [ISBN: 9789290832652], [hep-ph/0510198v1].

[62] A. Buckley et al., *Systematic event generator tuning for the LHC*. Eur. Phys. J. **C65**, (2009) 331–357, [doi: 10.1140/epjc/s10052-009-1196-7], [arXiv:0907.2973].

[63] G. Aad et al., the **ATLAS** Collaboration, *ATLAS Monte Carlo tunes for MC09*. ATL-PHYS-PUB 2010-002 (2010).

[64] P.Z. Skands, *The Perugia Tunes*. In *Proceedings of the First International Workshop on Multiple Partonic Interactions at the LHC (MPI@LHC08), Perugia, Italy, October 27-31, 2008*, edited by P. Bartalini and L. Fanò, 284–297 (DESY-PROC-2009-06, DESY, Germany) (2010), [arXiv:1003.4220].

[65] A. Moraes, C. Buttar, and I. Dawson, *Prediction for minimum bias and the underlying event at LHC energies*. Eur. Phys. J. **C50**, (2007) 435–466, [doi: 10.1140/epjc/s10052-007-0239-1].

[66] A. Moraes, C. Buttar, and I. Dawson, *Comparison of predictions for minimum bias event generators and consequences for ATLAS radiation background*. ATL-PHYS 2003-020 (2002).

[67] S.D. Ellis et al., *Jets in Hadron-Hadron Collisions*. Prog. Part. Nucl. Phys. **60**, (2008) 484–551, [doi: 10.1016/j.ppnp.2007.12.002], [arXiv:0712.2447].

[68] B. Flaugher and K. Meier, *A Compilation of Jet Finding Algorithms*. In *Proceedings of the 5th DPF Summer Study on High-energy Physics, Snowmass, CO, USA, 25 June - 23 July 1990*, edited by E.L. Berger, 134–136 (World Scientific, River Edge, NJ) (1992), [ISBN: 9789810209315].

[69] D. Krohn, J. Thaler, and L.-T. Wang, *Jets with Variable R*. JHEP **06**, (2009) 059, [doi: 10.1088/1126-6708/2009/06/059], [arXiv:0903.0392].

[70] T. Aaltonen et al., the **CDF** Collaboration, *Measurement of the Top Quark Mass and p anti-p → tt-bar Cross Section in the All-Hadronic Mode with the CDF II Detector*. Phys. Rev. **D81**, (2010) 052001, [doi: 10.1103/PhysRevD.81.052011], [arXiv:1002.0365].

[71] V.M. Abazov et al., the **DZero** Collaboration, *Measurement of the inclusive jet cross section in $p\bar{p}$ collisions at $\sqrt{s}$=1.96 TeV*. Phys. Rev. Lett. **101**, (2008) 062001, [doi: 10.1103/PhysRevLett.101.062001], [arXiv:0802.2400].

[72] J.E. Huth et al., *Toward a Standardization of Jet Definitions*. In *Proceedings of the 5th DPF Summer Study on High-energy Physics, Snowmass, CO, USA, 25 June - 23 July 1990*, edited by E.L. Berger (World Scientific, River Edge, NJ) (1992), [ISBN: 9789810209315].

[73] E.W.N. Glover and D.A. Kosower, *Recombination methods for jets in $p\bar{p}$ collisions*. Phys. Lett. **B367**, (1996) 369–376, [doi: 10.1016/0370-2693(95)01496-9], [hep-ph/9510420].

[74] G.C. Blazey et al., *Run II Jet Physics: Proceedings of the Run II QCD and Weak Boson Physics Workshop*. FERMILAB-Conf 00/092-E (2000), [hep-ex/0005012v2].

[75] D. Acosta et al., the **CMS** Collaboration, *CMS physics: Technical Design Report*. CERN-LHCC 2006-001 (2006), [ISBN: 9789290832683].

[76] A. Bhatti et al., the **CMS** Collaboration, *Performance of Jet Algorithms in CMS*. CMS PAS JME 07-003 (2009).

[77] G.P. Salam and G. Soyez, *A practical Seedless Infrared-Safe Cone jet algorithm*. JHEP **0705**, (2007) 086, [doi: 10.1088/1126-6708/2007/05/086], [arXiv:0704.0292].

[78] S.D. Ellis and D. Soper, *Successive combination jet algorithm for hadron collisions*. Phys. Rev. **D48**, (1993) 3160, [doi: 10.1103/PhysRevD.48.3160], [hep-ph/9305266].

[79] S. Catani, Yu.L. Dokshitzer, and B.R. Webber, *The $k_\perp$-clustering algorithm for jets in deep inelastic scattering and hadron collisions*. Phys. Lett. **B285**, (1992) 291–299, [doi: 10.1016/0370-2693(92)91467-N].

[80] S. Catani, Yu.L. Dokshitzer, B.R. Webber, and M.H. Seymour, *Longitudinally-invariant $k_\perp$-clustering algorithms for hadron-hadron collisions*. Nucl. Phys. **B406**, (1993) 187–224, [doi: 10.1016/0550-3213(93)90166-M].

[81] M. Cacciari, G.P. Salam, and G. Soyez, *The anti-k_t jet clustering algorithm*. JHEP **0804**, (2008) 063, [doi: 10.1088/1126-6708/2008/04/063], [arXiv:0802.1189].

[82] G. Choudalakis and M. Spousta, *Characteristic dip in average transverse energy density and multiplicity around the axis of Anti-kT jets*. ATL-COM-PHYS 2010-139 (2010).

[83] Yu.L. Dokshitzer, G.D. Leder, S. Moretti, and B.R. Webber, *Better Jet Clustering Algorithms*. JHEP **9708**, (1997) 001, [doi: 10.1088/1126-6708/1997/08/001], [hep-ph/9707323].

[84] M. Wobisch and T. Wengler, *Hadronization Corrections to Jet Cross Sections in Deep-Inelastic Scattering*. In *Proceedings of the Workshop on Monte Carlo Generators for HERA Physics*, edited by A.T. Doyle, G. Grindhammer, G. Ingelman, and H. Jung (DESY-PROC-1999-02, DESY, Germany) (1999), [hep-ph/9907280].

[85] L. Asquith et al., *Performance of Jet Algorithms in the ATLAS Detector*. ATL-COM-PHYS 2009-630 (2009).

[86] Z. Marshall, E.W. Hughes, and A. Schwartzman, *Track-based improvements in the jet energy resolution for ATLAS*. ATL-PHYS-INT 2008-029 (2008).

[87] G. Aad et al., the **ATLAS** Collaboration, *Expected Performance of the ATLAS Experiment, Detector, Trigger and Physics*. CERN-OPEN 2008-020, Geneva (2008), [ISBN: 9789290833215].

[88] C. Albajar et al., the **UA1** Collaboration, *Production of low transverse energy clusters in $p\bar{p}$ collisions at $\sqrt{s}$=0.2-0.9 TeV and their interpretation in terms of QCD jets*. Nucl. Phys. **B309**, (1988) 405, [doi: 10.1016/0550-3213(88)90450-6].

[89] S.D. Ellis, Z. Kunszt, and D.E. Soper, *Jets at hadron colliders at order α_s^3: A look inside*. Phys. Rev. Lett. **69**, (1992) 3615, [doi: 10.1103/PhysRevLett.69.3615].

[90] T. Kluge, *Measurements of Event and Jet Shapes at HERA*. Eur. Phys. J. **C33**, (2004) S416–S418, [doi: 10.1140/epjcd/s2004-03-1654-0], [hep-ex/0310063v1].

[91] J. Breitweg et al., the **ZEUS** Collaboration, *Measurement of jet shapes in high-Q^2 deep inelastic scattering at HERA*. Eur. Phys. J. **C8**, (1999) 367–380, [doi: 10.1007/s100529901107].

[92] C. Adloff et al., the **H1** Collaboration, *Measurement of internal jet structure in dijet production in deep-inelastic scattering at HERA*. Nucl. Phys. **B545**, (1999) 3–20, [doi: 10.1016/S0550-3213(99)00118-2].

[93] G.S. Abrams et al., the **Mark II** Collaboration, *Measurement of charged-particle inclusive distributions in hadronic decays of the Z boson*. Phys. Rev. Lett. **64**, (1990) 1334, [doi: 10.1103/PhysRevLett.64.1334].

[94] Y.K. Li et al., the **Amy** Collaboration, *Multihadron-event properties in e^+e^- annihilation at $\sqrt{s}$=52-57 GeV*. Phys. Rev. **D41**, (1990) 2675, [doi: 10.1103/PhysRevD.41.2675].

[95] R. Akers et al., the **OPAL** Collaboration, *QCD studies using a cone-based jet finding algorithm for e^+e^- collisions at LEP*. Zeit. f. Phys. **C63**, (1994) 197–211, [doi: 10.1007/BF01411011].

[96] D. Acosta et al., the **CDF** Collaboration, *Study of Jet Shapes in Inclusive Jet Production in p anti-p Collisions at s**(1/2) = 1.96 TeV*. Phys. Rev. **D71**, (2005) 112002, [doi: 10.1103/PhysRevD.71.112002], [hep-ex/0505013].

[97] F. Abe et al., the **CDF** Collaboration, *Measurement of jet shapes in p anti-p collisions at s**(1/2) = 1.8 TeV*. Phys. Rev. Lett. **70**, (1993) 713, [doi: 10.1103/PhysRevLett.70.713].

[98] S. Abachi et al., the **DZero** Collaboration, *Transverse Energy Distributions within Jets in p anti-p Collisions at s**(1/2) = 1.8 TeV*. Phys. Lett. **B357**, (1995) 500, [doi: 10.1016/0370-2693(95)00889-S].

[99] L.G. Almeida et al., *Top Jets at the LHC*. Phys. Rev. **D79**, (2009) 074012, [doi: 10.1103/PhysRevD.79.074012], [arXiv:0810.0934].

[100] I. Csabai, F. Czakó, and Z. Fodor, *Combined Neural Network-QCD classifier for quark and gluon jet separation*. Nucl. Phys. **B374**, (1992) 288–308, [doi: 10.1016/0550-3213(92)90354-E].

[101] J. Pumplin, *Variables for distinguishing between quark jets and gluon jets*. In *Proceedings of the 5th DPF Summer Study on High-energy Physics, Snowmass, CO, USA,*

25 June - 23 July 1990, edited by E.L. Berger, 174–178 (World Scientific, River Edge, NJ) (1992), [ISBN: 9789810209315].

[102] S.D. Ellis et al., *Jet Shapes and Jet Algorithms in SCET*. UCB-PTH 10/01 (2010), [arXiv:1001.0014].

[103] L.G. Almeida et al., *Substructure of high-p_T Jets at the LHC*. Phys. Rev. **D79**, (2009) 074017, [doi: 10.1103/PhysRevD.79.074017], [arXiv:0807.0234].

[104] M.H. Seymour, *The Subjet Multiplicity in Quark and Gluon Jets*. Phys. Lett. **B378**, (1996) 279–286, [doi: 10.1016/0370-2693(96)00399-1], [hep-ph/9603281v1].

[105] J. Pumplin, *Quark-Gluon Jet Differences at LEP*. Phys. Rev. **D48**, (1993) 1112–1116, [doi: 10.1103/PhysRevD.48.1112], [hep-ph/9301215].

[106] R.A. Briere et al., the **CLEO** Collaboration, *Comparison of Particle Production in Quark and Gluon Fragmentation at sqrt s $\sim$ 10 GeV*. Phys. Rev. **D76**, (2007) 012005, [doi: 10.1103/PhysRevD.76.012005], [arXiv:0704.2766].

[107] G. Abbiendi et al., the **OPAL** Collaboration, *Experimental properties of gluon and quark jets from a point source*. Eur. Phys. J. **C11**, (1999) 217–238, [doi: 10.1007/s100529900181], [hep-ex/9903027v1].

[108] L. Lönnblad, C. Peterson, and T. Rögnvaldsson, *Finding gluon jets with a neural trigger*. Phys. Rev. Lett. **65**, (1990) 1321–1324, [doi: 10.1103/PhysRevLett.65.1321].

[109] L. Jones, *Towards a systematic jet classification*. Phys. Rev. **D42**, (1990) 811–814, [doi: 10.1103/PhysRevD.42.811].

[110] T. Aaltonen et al., the **CDF** Collaboration, *Measurement of b-jet Shapes in Inclusive Jet Production in p anti-p Collisions at s**(1/2)=1.96 TeV*. Phys. Rev. **D78**, (2008) 072005, [doi: 10.1103/PhysRevD.78.072005].

[111] S. Frixione, M.L. Mangano, P. Nason, and G. Ridolfi, *Heavy-Quark Production*. In *Heavy Flavours II*, edited by A.J. Buras and M. Lindner, 609–706 (World Scientific, Singapore) (1998), [ISBN: 9810222157], [hep-ph/9702287].

[112] M. Martišíková for the H1 Collaboration, *Study of Jet Shapes in Charm Photoproduction at HERA*. AIP Conf. Proc. **792**, (2005) 819–822, [doi: 10.1063/1.2122162].

[113] J. Gallicchio and M.D. Schwartz, *Seeing in Color: Jet Superstructure*. Phys. Rev. Lett. **105**, (2010) 022001, [doi: 10.1103/PhysRevLett.105.022001], [arXiv:1001.5027].

[114] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, *Jet substructure as a new Higgs search channel at the LHC*. Phys. Rev. Lett. **100**, (2008) 242001, [doi: 10.1103/PhysRevLett.100.242001], [arXiv:0802.2470].

[115] S.D. Ellis, C.K. Vermilion, and J.R. Walsh, *Techniques for improved heavy particle searches with jet substructure*. Phys. Rev. **D80**, (2009) 051501, [doi: 10.1103/PhysRevD.80.051501], [arXiv:0903.5081].

[116] D. Krohn, J. Thaler, and L.-T. Wang, *Jet Trimming*. JHEP **2010**, (2009) 084, [doi: 10.1007/JHEP02(2010)084], [arXiv:0912.1342].

[117] M. Vázquez, *Substructure of Jets at HERA*. In *PHOTON 2001: Proceedings of the International Conference on the Structure and Interactions of the Photon Including the 14th International Workshop on Photon-Photon Collisions*, edited by M.-N. Kienzle-Focacci and M. Wadhwa, 75–79 (World Scientific, River Edge, NJ) (2001), [ISBN: 9789812380531], [doi: 10.1142/9789812777157_0015], [hep-ex/0110085].

[118] O.S. Bruening et al. (eds.) *LHC Design Report*, volume v.1: The LHC Main Ring (CERN, Geneva) (2004), [ISBN: 9789290832249], also CERN-2004-003-V-1.

[119] O.S. Bruening et al. (eds.) *LHC Design Report*, volume v.2: The LHC Infrastructure and General Services (CERN, Geneva) (2004), [ISBN: 9789290832249], also CERN-2004-003-V-2.

[120] M. Benedikt et al. (eds.) *LHC Design Report*, volume v.3: The LHC Injector Chain (CERN, Geneva) (2004), [ISBN: 9789290832393], also CERN-2004-003-V-3.

[121] The LEP Injector Study Group, *LEP design report, Vol. 1, The LEP injector chain*. CERN-LEP-TH 83-29, Geneva (1983).

[122] The LEP Injector Study Group, *LEP design report, Vol. 2, The LEP main ring*. CERN-LEP 84-01, Geneva (1984).

[123] C. Rubbia et al., *Report of the Long-range Planning Committee to the CERN Council: 83rd session of Council, Geneva, 19 June 1987*. CERN 1658, Geneva (1987).

[124] G. Jackson and G.W. Foster, *Storage ring for enhanced antiproton production at Fermilab*. In *Proceedings of the 16th IEEE Particle Accelerator Conference, Dallas, TX, USA, 1-5 May 1995*, volume 1, 473–475 (IEEE, Piscataway, NJ) (1995), [ISBN: 9780780329348], [doi: 10.1109/PAC.1995.504695].

[125] G. Aad et al., the **ATLAS** Collaboration, *Characterization of Interaction-Point Beam Parameters Using the pp Event-Vertex Distribution Reconstructed in the ATLAS Detector at the LHC*. ATLAS-CONF 2010-027 (2010).

[126] A. Breskin and R. Voss (eds.) *The CERN Large Hadron Collider: Accelerator and Experiments*, volume Volume 1: LHC Machine, ALICE, and ATLAS (CERN, Geneva) (2009), [ISBN: 9789290833376].

[127] A. Breskin and R. Voss (eds.) *The CERN Large Hadron Collider: Accelerator and Experiments*, volume Volume 2: CMS, LHCb, LHCf, and TOTEM (CERN, Geneva) (2009), [ISBN: 9789290833383].

[128] G. Aad et al., the **ATLAS** Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*. JINST **3**, (2008) S08003, [doi: 10.1088/1748-0221/3/08/S08003].

[129] O. Adriani et al., the **LHCf** Collaboration, *The LHCf Detector at the CERN Large Hadron Collider*. JINST **3**, (2008) S08006, [doi: 10.1088/1748-0221/3/08/S08006].

[130] K. Aamodt et al., the **ALICE** Collaboration, *The ALICE Experiment at the CERN LHC*. JINST **3**, (2008) S08002, [doi: 10.1088/1748-0221/3/08/S08002].

[131] C. Chatrchyan et al., the **CMS** Collaboration, *The CMS Experiment at the CERN LHC*. JINST **3**, (2008) S08004, [doi: 10.1088/1748-0221/3/08/S08004].

[132] G. Anelli et al., the **TOTEM** Collaboration, *The TOTEM Experiment at the CERN Large Hadron Collider*. JINST **3**, (2008) S08007, [doi: 10.1088/1748-0221/3/08/S08007].

[133] A.A. Alves Jr et al., the **LHCb** Collaboration, *The LHCb Detector at the CERN LHC*. JINST **3**, (2008) S08005, [doi: 10.1088/1748-0221/3/08/S08005].

[134] Steve Myers, *LEP Operation*. Talk, from the Proton Synchrotron to the Large Hadron Collider - 50 Years of Nobel Memories in High-Energy Physics. Video online.

[135] A. Verweij, *Training the dipoles*. In *Proceedings of the Chamonix 2009 Workshop on LHC Performance, Chamonix, France, 2-6 Feb 2009*, edited by C. Carli, 25–28 (CERN, Geneva) (2009).

[136] I. Bird et al., *LHC computing Grid. Technical design report*. CERN-LHCC 2005-024 (2005), [ISBN: 9789290832539].

[137] G. Aad et al., *ATLAS pixel detector electronics and sensors*. JINST **3**, (2008) 07007, [doi: 10.1088/1748-0221/3/07/P07007].

[138] M. Marcisovsky, *ATLAS Insertable B-Layer*. In *Proceedings of the 11th International Workshop On Radiation Imaging Detectors, Prague, Czech Republic, 28 Jun - 02 Jul 2009 (iWoRiD 2009)* (2009), to appear. Now ATL-INDET-PROC 2009-012.

[139] The ATLAS Inner Detector Community, the **ATLAS** Collaboration, *ATLAS Inner Detector Technical Design Report*. CERN/LHCC 97-16 (1997).

[140] G. Aad et al., the **ATLAS** Collaboration, *Performance of primary vertex reconstruction in proton-proton collisions at sqrt(s)=7 TeV in the ATLAS experiment*. ATLAS-CONF 2010-069 (2010).

[141] G. Aad et al., the **ATLAS** Collaboration, *Charged Particle Multiplicities in pp interactions at $\sqrt{s}$=7 TeV*. ATLAS-CONF 2010-048 (2010).

[142] G. Aad et al., the **ATLAS** Collaboration, *Readiness of the ATLAS Liquid Argon Calorimeter for LHC Collisions*. ATLAS-LAR 2009-01-004 (2009), accepted for publication in Eur. Phys. J. C, [arXiv:0912.2642].

[143] M. Aharrouche et al., *Energy linearity and resolution of the ATLAS electromagnetic barrel calorimeter in an electron test-beam*. Nucl. Inst. Meth. **A568**, (2006) 601–623, [doi: 10.1016/j.nima.2006.07.053].

[144] P. Mermod, G. Arabidze, D. Milstead, and R. Stanek, *Effects of ATLAS Tile calorimeter failures on jets and missing transverse energy measurements*. ATL-TILECAL-PUB 2008-011-1 (2008).

[145] G. Choudalakis and E. Feng, *Use of Tracks to Reconstruct Jets in Damaged Calorimeter Regions*. ATL-PHYS-INT 2009-017 (2009).

[146] P. Adragna et al., *Measurement of Pion and Proton Longitudinal Shower Profiles up to 20 Nuclear Interaction Lengths with the ATLAS Tile Calorimeter*. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **615**, (2010) 158–181, [doi: 10.1016/j.nima.2010.01.037].

[147] P. Adragna et al., *Testbeam Studies of Production Modules of the ATLAS Tile Calorimeter*. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **606**, (2009) 362 – 394, [doi: 10.1016/j.nima.2009.04.009].

[148] E. Abat et al., *Response and Shower Topology of 2 to 180 GeV Pions Measured with the ATLAS Barrel Calorimeter at the CERN Test-beam and Comparison to Monte Carlo Simulations*. ATL-CAL-PUB 2010-001 (2010).

[149] G. Aad et al., the **ATLAS** Collaboration, *Commissioning of the ATLAS Muon Spectrometer with Cosmic Rays*. ATLAS-MUON 2010-01-004 (2010), submitted to Eur. Phys. J. C, [arXiv:1006.4384].

[150] S. Baranov et al., *ATLAS Radiation Background Task Force Summary Document*. ATL-GEN 2005-001 (2005).

[151] G. Aad et al., the **ATLAS** Collaboration, *Luminosity Determination Using the ATLAS Detector*. ATLAS-CONF 2010-060 (2010).

[152] G. Aad et al., the **ATLAS** Collaboration, *The ATLAS Monte Carlo Project*. ATLAS-SOFT 2010-01-004 (2010), accepted for publication in Eur. Phys. J. C, [arXiv:1005.4568].

[153] S. Agostinelli et al., *Geant4 - a simulation toolkit*. Nucl. Inst. Meth. **A506**, (2003) 250–303, [doi: 10.1016/S0168-9002(03)01368-8].

[154] J. Allison et al., *Geant4 Developments and Applications*. IEEE Transactions on Nuclear Science **53**, (2006) 270–278, [doi: 10.1109/TNS.2006.869826].

[155] A. Rimoldi et al., *First Report of the Simulation Optimization Group*. ATL-SOFT-PUB 2008-002 (2008).

[156] A. Rimoldi et al., *Final Report of the Simulation Optimization Task Force*. ATL-SOFT-PUB 2008-004 (2008).

[157] J. Adelman et al., *ATLAS Offline Data Quality Monitoring*. In *Proceedings of the 17th International Conference on Computing in High Energy and Nuclear Physics, Prague, Czech Republic, 21-27 March 2009*, 042018 (IOP Journal of Physics: Conference Series) (2010), [doi: 10.1088/1742-6596/219/4/042018].

[158] M. Hauschild, B. Heinemann, A. Hoecker, and P. Onyisi, *Data Quality Efficiency for Periods A-D*. ATL-COM-PHYS 2010-577 (2010).

[159] G. Aad et al., the **ATLAS** Collaboration, *Data-Quality Requirements and Event Cleaning for Jets and Missing Transverse Energy Reconstruction with the ATLAS Detector in Proton-Proton Collisions at a Center-of-Mass Energy of sqrt(s)=7 TeV*. ATLAS-CONF 2010-038 (2010).

[160] G. Aad et al., the **ATLAS** Collaboration, *Performance of the ATLAS Jet Trigger in the Early $\sqrt{s}$=7 TeV Data*. ATLAS-CONF 2010-047 (2010).

[161] M. Cacciari and G.P. Salam, *Dispelling the N**3 myth for the k(t) jet-finder*. Phys. Lett. **B641**, (2006) 57–61, [doi: 10.1016/j.physletb.2006.08.037], [hep-ph/0512210].

[162] G. Aad et al., the **ATLAS** Collaboration, *ATLAS Computing Technical Design Report*. CERN-LHCC 2005-022 (2005), see also the Athena Online Manual.

[163] J.M. Butterworth, J.R. Forshaw, and M.H. Seymour, *Multiparton interactions in photoproduction at HERA*. Zeit. f. Phys. **C72**, (1996) 637–646, [doi: 10.1007/s002880050286], [hep-ph/9601371].

[164] R. Achenbach et al., *The ATLAS Level-1 Calorimeter Trigger*. ATL-DAQ-PUB 2008-001 (2008).

[165] G. Aad et al., the **ATLAS** Collaboration, *Performance of the ATLAS jet trigger with pp collisions at $\sqrt{s}$ = 900 GeV*. ATLAS-CONF 2010-028 (2010).

[166] M. Begel and M. Campanelli, *Jet Trigger Plots for ICHEP*. ATL-COM-DAQ 2010-086 (2010).

[167] A.I. Drozhdin, M. Huhtinen, and N.V. Mokhov, *Accelerator related backgrounds in the CMS detector at LHC*. Nucl. Instr. Meth. **A381**, (2004) 531–544, [doi:10.1016/S0168-9002(96)00807-8].

[168] A. Rossi, *Residual Gas Density Estimations in the LHC Insertion Regions IR1 and IR5 and the Experimental Regions of ATLAS and CMS for Different Beam Operations*. LHC Project Report 783 (2004).

[169] H. Okawa et al., *Calorimeter Commissioning with Cosmic Rays: Cells, Clusters, Jets, and Missing E_T*. ATL-COM-CAL 2008-007 (2009).

[170] G. Aad et al., the **ATLAS** Collaboration, *Jet energy scale and its systematic uncertainty in ATLAS for jets produced in proton-proton collisions at $\sqrt{s}$=7 TeV*. ATLAS-CONF 2010-056 (2010).

[171] G. Aad et al., the **ATLAS** Collaboration, *Properties of Jets and Inputs to Jet Reconstruction and Calibration with the ATLAS Detector Using Proton-Proton Collisions at $\sqrt{s}$=7 TeV*. ATLAS-CONF 2010-053 (2010).

[172] G. Aad et al., the **ATLAS** Collaboration, *Measurement of jet production in proton-proton collisions at 7 TeV centre-of-mass energy with the ATLAS Detector*. ATLAS-CONF 2010-050 (2010).

[173] G. Aad et al., the **ATLAS** Collaboration, *Measurement of inclusive jet and dijet cross sections in proton-proton collisions at 7 TeV centre-of-mass energy with the ATLAS detector* [arXiv:1009.5908], submitted to Eur. Phys. J. C.

[174] G. Aad et al., the **ATLAS** Collaboration, *Jet energy resolution and reconstruction efficiencies from in-situ techniques with the ATLAS Detector Using Proton-Proton Collisions at a Center of Mass Energy $\sqrt{s}$=7 TeV*. ATLAS-CONF 2010-054 (2010).

[175] G. Cowan, *A Survey of Unfolding Methods for Particle Physics*. In *Proceedings of Advanced Statistical Techniques in Particle Physics, Durham, England, 18-22 March 2002*, edited by M.R. Whalley and L. Lyons, 248–257 (2002).

[176] G. Folger and J.-P. Wellisch, *String Parton Models in Geant4*. In *Proceedings of the 2003 Conference for Computing in High Energy and Nuclear Physics, La Jolla, CA, USA, 24-28 March 2003*, MOMT007 (2003), [nucl-th/0306007].

[177] N.S. Amelin et al., *Transverse flow and collectivity in ultrarelativistic heavy-ion collisions*. Phys. Rev. Lett. **67**, (1991) 1523–1526, [doi: 10.1103/PhysRevLett.67.1523].

[178] L.V. Bravina, *Scaling violation of transverse flow in heavy ion collisions at AGS energies*. Phys. Lett. **B344**, (1995) 49–54, [doi: 10.1016/0370-2693(94)01560-Y].

[179] V. Lara and J.-P. Wellisch, *Pre-equilibrium and equilibrium decays in Geant4*. In *Proceedings of the CHEP 2000 Conference, Padova, Italy, 7-11 February 2000*, edited by M. Mazzucato, 52–55 (2000).

[180] M.P. Guthrie, R.G. Alsmiller Jr., and H.W. Bertini, *Calculation of the capture of negative pions in light elements and comparison with experiments pertaining to cancer radiotherapy*. Nucl. Instrum. and Methods **66**, (1968) 29–36, [doi: 10.1016/0029-554X(68)90054-2].

[181] A. Heikkinen, N. Stepanov, and J.-P. Wellisch, *Bertini Intra-nuclear cascade Implementation in Geant4*. In *Proceedings of the 2003 Conference for Computing in High Energy and Nuclear Physics, La Jolla, CA, USA, 24-28 March 2003*, MOMT008 (2003), [nucl-th/0306008].

[182] A. Ribon et al., *Status of* GEANT4 *hadronic physics for the simulation of LHC experiments at the start of LHC physics program*. CERN-LCGAPP 2010-01 (2010).

[183] B. Andersson, G. Gustafson, and B. Nilsson-Almqvist, *A model for low-pT hadronic reactions with generalizations to hadron-nucleus and nucleus-nucleus collisions*. Nucl. Phys. **B281**, (1987) 289–309, [doi: 10.1016/0550-3213(87)90257-4].

[184] B. Nillson-Almquist and E. Stenlund, *Interactions between hadrons and nuclei: The Lund Monte Carlo - FRITIOF version 1.6*. Comp. Phys. Commun. **43**, (1987) 387–397, [doi: 10.1016/0010-4655(87)90056-7].

[185] G. Aad et al., the **ATLAS** Collaboration, *Probing the material in front of the ATLAS electromagnetic calorimeter with energy flow from* $\sqrt{s}$*=7 TeV minimum bias events*. ATLAS-CONF 2010-037 (2010).

[186] G. Aad et al., the **ATLAS** Collaboration, *Study of the Material Budget in the ATLAS Inner Detector with Kshort decays in collision data at sqrt(s)=900 GeV*. ATL-CONF 2010-019 (2010).

[187] M. Begel et al., *Azimuthal Decorrelations in Dijet Events at sqrt{s}=7 TeV*. ATL-COM-PHYS 2010-564 (2010), to become an ATLAS CONF note.

[188] D.W. Miller and A. Schwartzman, *Jet energy pile-up subtraction for mutliple luminosity scenarios at the LHC*. ATL-COM-PHYS 2010-141 (2010).

[189] G. Aad et al., the **ATLAS** Collaboration, *Charged particle multiplicities in pp interactions at $\sqrt{s}$=0.9 and 7 TeV in diffractive limited phase space measured with the ATLAS detector at the LHC and a new* PYTHIA*6 tune*. ATLAS-CONF 2010-031 (2010).

[190] P.Z. Skands, *The Perugia Tunes*. MCnet 10/08 (2010), [arXiv:1005.3457].

[191] P.Z. Skands, *Some Interesting Min-Bias Distributions for Early LHC Runs*. In *Proceedings of the 5th Les Houches Workshop on Physics at TeV Colliders, Les Houches, France, 11-29 June 2007* (IN2P3, Paris) (2007).

[192] B. Demirkoz, M. Martinez, and F. Vives, *First Measurement of Jet Shapes in proton-proton collisions at 7 TeV centre-of-mass energy with the ATLAS Detector*. ATLAS-COM-PHYS 2010-561 (2010), to become an ATLAS CONF note.

[193] C. Chatrchyan et al., the **CMS** Collaboration, *Jet Transverse Structure and Momentum Distribution in pp Collisions at 7 TeV*. CMS-PAS-QCD 10-014 (2010).

[194] G. Aad et al., the **ATLAS** Collaboration, *Track-based underlying event measurements in pp collisions at sqrt{s}=900 GeV and 7 TeV with the ATLAS Detector at the LHC* (2010-081).

[195] G. Aad et al., the **ATLAS** Collaboration, *ATLAS Calorimeter Response to Single Isolated Hadrons and Estimation of the Calorimeter Jet Scale Uncertainty* (2010-052).

[196] G. Aad et al., the **ATLAS** Collaboration, *Response of the ATLAS calorimeter to single isolated hadrons produced in proton proton collisions* (2010-017).

[197] G. Aad et al., the **ATLAS** Collaboration, *Search for New Particles in Two-Jet Final States in 7 TeV Proton-Proton Collisions with the ATLAS Detector at the LHC*. Phys. Rev. Lett. **105**, (2010) 161801, [arXiv:1008.2461].

[198] G. Aad et al., the **ATLAS** Collaboration, *Search for Quark Contact Interactions in Dijet Angular Distributions in 7 TeV Proton-Proton Collisions with the ATLAS Detector at the LHC* [arXiv:1009.5069], submitted to Phys. Lett. B.

[199] G. Aad et al., the **ATLAS** Collaboration, *Search for top pair candidate events in ATLAS at* $\sqrt{s}$*=7 TeV* (2010-063).

[200] David Lopez. Private communication. The quark-gluon/jet association algorithm was the same one as was used in this thesis.

[201] D. Lopez-Mateos, E.W. Hughes, and A. Schwartzman, *Reduction of the Flavor Dependence of the Jet Energy Scale Using a Correction that Depends on the Transverse Structure of the Jet*. ATL-PHYS-INT 2010-046 (2010).

[202] G. Choudalakis, E. Feng, A. Messina, and L. Pribyl, *In Situ Jet Energy Scale Calibration Using Direct* γ*+jet Balance*. ATL-PHYS-INT 2010-023 (2010).

[203] S. Frixione, P. Nason, and B.R. Webber, *Matching NLO QCD and parton showers in heavy flavour production*. JHEP **08**, (2003) 007, [hep-ph/0305252].

[204] N. Kidonakis, *Top quark pair and single top production at Tevatron and LHC energies*. In *Proceedings of the International Conference on High Energy Physics (ICHEP2010), Paris, France, 22-28 July 2010* (2010), to appear, [arXiv:1008.2460].

[205] C. Amsler et al., the **Particle Data Group** Collaboration. Physics Letters **B667**, (2008) 1.

[206] S.P. Martin, *A Supersymmetry Primer* [hep-ph/9709356].

[207] T. Affolder et al., the **CDF** Collaboration, *Measurement of the Strong Coupling Constant from Inclusive Jet Production at the Tevatron* $\bar{p}p$ *Collider*. Phys. Rev. Lett. **88**, (2001) 042001, [doi: 10.1103/PhysRevLett.88.042001], [hep-ex/0108034].

[208] A. Höcker and V. Kartvelishvili, *SVD approach to data unfolding*. Nucl. Instrum. and Methods **A372**, (1996) 469–481, [doi: 10.1016/0168-9002(95)01478-0].

[209] G. D'Agostini, *A multidimensional unfolding method based on Bayes' theorem*. Nucl. Instrum. and Methods **A362**, (1995) 487–498, [doi: 10.1016/0168-9002(95)00274-X].

[210] H.M. Gray et al., *Aspects of an Inclusive Search for Supersymmetry in the 0-lepton channel at sqrt(s) = 10 TeV*. ATL-COM-PHYS 2009-622 (2009).